

Tech Challenge

Aluno: Guilherme Neto Tonini

Matrícula: RM353862

Turma: Pós Tech – 4DTAT

Resumo

Este projeto tem como objetivo apresentar uma análise concisa dos dados do IBGE coletados durante a pandemia de COVID-19, abrangendo questionários aplicados a uma amostra da população brasileira ao longo de três meses: agosto, setembro e outubro de 2020. Para o desenvolvimento, foi escolhida a plataforma Google Colab, que oferece armazenamento em nuvem, uma versão gratuita, suporte ao Jupyter Notebook e integração com ferramentas como as bibliotecas pandas e Spark. O estudo se desdobra em três análises principais. A primeira foca no comportamento social, investigando a adoção de restrições, a realização de testes de COVID-19 e a ocorrência de sintomas entre os entrevistados. A segunda análise explora o comportamento dos entrevistados em relação à busca por atendimento médico, diferenciando entre os serviços público e privado, considerando a posse de plano de saúde como um fator econômico. A terceira análise concentra-se em aspectos clínicos, observando a necessidade de internações hospitalares e sua relação com diversos diagnósticos, além de examinar a necessidade de aparelhos de respiração artificial em pacientes que testaram positivo para COVID-19. Além dessas análises, o projeto sugere duas ações recomendadas para futuros surtos de COVID-19. A primeira, fundamentada na análise 3.1, propõe a ampliação do acesso a testes de COVID, com foco especial em pessoas que não adotam restrições, apresentam sintomas e procuram atendimento. A segunda, baseada na análise 3.3, sugere um monitoramento mais rigoroso e cuidados preventivos para grupos de risco, como aqueles com hipertensão ou doenças cardíacas, identificados como mais propensos à internação.

Palavras-Chave: COVID-19, Análise de dados, IBGE

Sumário

1	Introdução	4
2	Pré-Análise.....	4
3	Análise.....	6
3.1	Análise Comportamental sobre Restrições.....	6
3.2	Análise Comportamental Sobre Atendimento Médico.....	7
3.3	Análise Comparativa de Internação e Aparelhos Respiratórios	9
4	Possíveis Ações	12

1 Introdução

Este projeto tem como objetivo apresentar uma breve análise referente a três meses — agosto, setembro e outubro de 2020 —, período que fez parte da pandemia de COVID-19. Os dados foram obtidos diretamente do site do IBGE e incluem uma série de questionamentos aplicados a uma amostra da população brasileira. A pesquisa abrange informações gerais, como escolaridade, raça, gênero e localização geográfica, além de dois questionários específicos.

O primeiro questionário aborda questões de saúde, como sintomas relacionados à COVID-19, busca por atendimento médico, medidas tomadas e o tipo de estabelecimento procurado. O segundo questionário foca em questões relacionadas ao trabalho, como ocupação e atividade, afastamento, trabalho remoto, busca por emprego, carga horária semanal, rendimentos, entre outros aspectos importantes.

Para aqueles que desejarem acessar o código completo utilizado neste trabalho, ele está disponível no link: [Google Colab - Código do Projeto](#).

2 Pré-Análise

Este estudo utiliza um banco de dados em nuvem, com a plataforma escolhida sendo o Google Colab, devido à sua versão gratuita, suporte a notebooks Jupyter, e integração com diversas ferramentas, como as utilizadas neste projeto: a biblioteca Pandas, da linguagem Python, e o uso de SQL através da ferramenta Apache Spark. O DataFrame inicial contém 145 colunas e mais de 1 milhão de linhas. No entanto, foi solicitado pelo Head de Dados que o número de variáveis fosse reduzido para 20 colunas. Para atender a essa solicitação, foram realizados diversos agrupamentos entre colunas relacionadas, como ilustrado na imagem abaixo, permitindo condensar uma maior quantidade de informações em menos colunas.

Embora esse processo de agrupamento possa tornar algumas pesquisas e interpretações mais desafiadoras, ele resulta em um DataFrame mais conciso e eficiente para análise, facilitando o tratamento e visualização de dados sem comprometer a abrangência do conteúdo.

Essa abordagem foi utilizada para garantir que as perguntas mais relevantes fossem preservadas no conjunto final de dados, mantendo o equilíbrio entre parcimônia e informação detalhada.

```

# Inicializa a nova coluna 'Sintomas' como uma string vazia
df_resumo['Sintomas'] = ''

# Verifica cada coluna e concatena o número correspondente se o valor for 1
df_resumo['Sintomas'] += np.where(df_resumo['B0011'] == '1', '1,', '')
df_resumo['Sintomas'] += np.where(df_resumo['B0012'] == '1', '2,', '')
df_resumo['Sintomas'] += np.where(df_resumo['B0013'] == '1', '3,', '')
df_resumo['Sintomas'] += np.where(df_resumo['B0014'] == '1', '4,', '')
df_resumo['Sintomas'] += np.where(df_resumo['B0015'] == '1', '5,', '')
df_resumo['Sintomas'] += np.where(df_resumo['B0016'] == '1', '6,', '')
df_resumo['Sintomas'] += np.where(df_resumo['B0017'] == '1', '7,', '')
df_resumo['Sintomas'] += np.where(df_resumo['B0018'] == '1', '8,', '')
df_resumo['Sintomas'] += np.where(df_resumo['B0019'] == '1', '9,', '')
df_resumo['Sintomas'] += np.where(df_resumo['B00110'] == '1', '10,', '')
df_resumo['Sintomas'] += np.where(df_resumo['B00111'] == '1', '11,', '')
df_resumo['Sintomas'] += np.where(df_resumo['B00112'] == '1', '12,', '')
df_resumo['Sintomas'] += np.where(df_resumo['B00113'] == '1', '13,', '')

# Se todos os valores das colunas de sintomas forem 2, preenche 'Sintomas' com 14
df_resumo['Sintomas'] = np.where(
    (df_resumo[['B0011', 'B0012', 'B0013', 'B0014', 'B0015', 'B0016', 'B0017', 'B0018', 'B0019',
                'B00110', 'B00111', 'B00112', 'B00113']] == '2').all(axis=1), '14', df_resumo['Sintomas']
)

# Remove a vírgula final e substitui strings vazias por 'Nao_Aplicavelo_Aplicavel'
df_resumo['Sintomas'] = df_resumo['Sintomas'].str.rstrip(',')
df_resumo['Sintomas'] = df_resumo['Sintomas'].replace('', 'Nao_Aplicavel')

```

A imagem exibe 13 colunas do DataFrame original, que correspondem aos sintomas relatados no questionário. Cada uma dessas colunas contém a resposta dos entrevistados, sendo que, se o valor for "1", significa que o entrevistado relatou ter sofrido esse sintoma na última semana. Nesse caso, o número associado ao sintoma é registrado na coluna correspondente. Além disso, há um código específico atribuído quando o entrevistado não relatou nenhum dos sintomas listados no questionário. Essa lógica foi aplicada a outras partes do DataFrame final.

O conjunto de dados apresenta um número significativo de valores NaN, que correspondem a células vazias. De acordo com o site do IBGE, essas lacunas indicam que a pergunta em questão não era aplicável ao entrevistado. Para lidar com essa situação os valores ausentes foram substituídos por "Nao_Aplicavel".

Por fim, algumas colunas foram mantidas em formato numérico devido ao fato de as respostas conterem um grande número de caracteres, o que poderia resultar no truncamento dessas informações durante as consultas e análises. Embora essa abordagem possa tornar a interpretação dos resultados um pouco mais trabalhosa, ela garante que não haja perda de dados importantes. Além disso, essa prática segue o próprio padrão adotado pelo IBGE.

Para facilitar a interpretação dos resultados, foram criados dicionários para cada coluna, associando os números às suas respectivas respostas descritivas. Esses dicionários permitem uma leitura mais clara e direta dos dados, mitigando a dificuldade trazida pelo uso de códigos numéricos e garantindo que os analistas possam compreender rapidamente os resultados durante a análise.

3 Análise

O estudo deste DataFrame foi dividido em três análises principais:

1. **Análise Comportamental sobre Restrições:** Esta análise aborda as restrições adotadas pelos entrevistados, a realização de testes de COVID-19 e os sintomas mais comuns entre pessoas que não algum tipo de restrição.
2. **Análise Comportamental sobre Atendimento Médico:** Nesta etapa, a análise foca na procura por atendimento entre os serviços públicos e privados, relacionando essas escolhas com um parâmetro econômico — a posse de plano de saúde.
3. **Análise Comparativa de Internação e Aparelhos Respiratórios:** A última análise compara pacientes que necessitaram de internação e uso de aparelhos respiratórios e que testaram positivo para COVID-19 com a porcentagem geral dos entrevistados. Além disso, foram analisados quais diagnósticos anteriores foram os mais comuns entre esses pacientes.

3.1 Análise Comportamental sobre Restrições

A pesquisa inicial consistiu em analisar a porcentagem de entrevistados que adotaram algum método de restrição. Como essa variável é mantida com um valor numérico, o dicionário correspondente é disponibilizado.

restricao	quantidade	porcentagem
3	477293	41.35
2	444729	38.53
4	190169	16.48
1	36100	3.13
9	5988	0.52

- 1 – Não fez restrição. Levou a vida normal como antes da pandemia
- 2 – Reduziu o contato com pessoas, mas continuou saindo de casa para trabalho ou atividades não essenciais e/ ou recebendo visitas
- 3-Ficou em casa e só saiu em caso de necessidade básica
- 4 - Ficou rigorosamente em casa
- 9 – Ignorado

A pesquisa revela que, entre os entrevistados, pouco mais de 36 mil indivíduos — o que representa aproximadamente 3% — não adotaram nenhum tipo de restrição. No próximo passo, será isolada a quantidade de entrevistados que não realizaram restrições, e a variável relacionada ao número de indivíduos que efetivamente realizaram testes de COVID-19 será introduzida.

teste	quantidade	porcentagem
Não	31102	86.16
Sim	4980	13.8
ignorado	18	0.05

Entre os entrevistados que não adotaram nenhuma restrição, aproximadamente 86% também não realizaram testes de COVID-19. No total, 31.102 indivíduos não adotaram qualquer forma de restrição e não realizaram testes, correspondendo a cerca de 2,69% de todo o dataset. Neste momento, é relevante analisar a porcentagem desse grupo que relata ter apresentado algum sintoma.

sintoma	total_quantidade	porcentagem
14	29721	95.56
5	493	1.59
8	481	1.55
2	394	1.27
3	328	1.05
12	245	0.79
1	206	0.66
9	184	0.59
13	123	0.4
Nao_Aplicavel	116	0.37
4	109	0.35
10	97	0.31
11	92	0.3
6	92	0.3
7	88	0.28

- 1 – Febre
- 2 – Tosse
- 3 – Dor de garganta
- 4 – Dificuldade para respirar
- 5 – Dor de cabeça
- 6 – Dor no peito
- 7 – Náusea
- 8 – Nariz entupido ou escorrendo
- 9 – Fadiga
- 10 – Dor nos olhos
- 11 – Perda de cheiro ou sabor
- 12 – Dor muscular
- 13 – Diarreia
- 14 - Nenhum

Entre os entrevistados que não adotaram nenhum tipo de restrição e não realizaram testes de COVID-19, aproximadamente 95,5% relataram não ter apresentado nenhum sintoma. Dos 4,5% que afirmaram ter sentido algum sintoma na última semana, os mais frequentes foram dor de cabeça, nariz entupido ou escorrendo e tosse.

3.2 Análise Comportamental Sobre Atendimento Médico

Para iniciar a análise, foi investigada a parcela da população que buscou atendimento médico, levando em consideração se o atendimento ocorreu em serviços públicos ou privados, com um filtro adicional relacionado à posse de plano de saúde. No entanto, a maioria dos dados apresenta a característica "Não Aplicável". Como o foco da pesquisa são os casos em que o atendimento efetivamente ocorreu e os valores são aplicáveis, esses dados foram isolados da análise atual, garantindo maior precisão nos resultados.

plano_de_saude	tipo_atendimento	quantidade_total
Não	nenhum	196
Não	público	9794
Não	privado	690
Sim	nenhum	94
Sim	privado	2084
Sim	público	1690
ignorado	público	3

A pesquisa revela que a maior parte da população corresponde a indivíduos sem plano de saúde, que buscaram atendimento em serviços públicos. Ao comparar esse grupo com aqueles que procuraram serviços privados, verifica-se que apenas 7% das pessoas sem plano de saúde buscaram atendimento no setor privado, em relação ao número de atendimentos no setor público. No entanto, essa diferença é significativamente menor entre os entrevistados que possuem plano de saúde, onde o número de pessoas que utilizaram o setor público representa 81% do total que buscou atendimento no setor privado. Além disso, cerca de 10% dos entrevistados recorreram a mais de uma forma de atendimento, alternando entre exclusivamente público, privado, ou uma combinação de ambos, mostrado na pesquisa abaixo.

plano_de_saude	tipo_servico	quantidade	porcentagem
Não	público-privado	136	10.63
Não	público	761	59.5
Não	privado	17	1.33
Sim	privado	84	6.57
Sim	público-privado	154	12.04
Sim	público	127	9.93

É interessante observar que, a maior parcela entre entrevistados que possuem plano de saúde e buscaram mais de uma forma de atendimento afirmaram que buscaram atendimento tanto no serviço público quanto no privado, além de apresentar a menor porcentagem desse grupo em atendimentos exclusivamente privados. Por outro lado, o grupo de pessoas sem plano de saúde que buscaram mais de uma forma de atendimento apresenta números bem discrepantes entre as opções, mas ainda assim apresentando valores para a opção de atendimentos exclusivamente no setor privado.

tipo_atendimento	quantidade_total	porcentagem_total
público	10142	79.39
privado	2343	18.34
nenhum	290	2.27

De forma geral, cerca de 79% dos atendimentos foram feitos no serviço público de saúde, enquanto pouco mais de 18% foram no serviço privado.

3.3 Análise Comparativa de Internação e Aparelhos Respiratórios

O estudo realizado dos três meses de pesquisa revelou que menos de 1% dos entrevistados afirmou ter necessitado de internação no período analisado, o que ainda corresponde a um número considerável de 634 pessoas. Dentro desse grupo, foi analisado quantos entrevistados testaram positivo para COVID-19.

teste_resultado	quantidade	porcentagem
Positivo	287	45.27
Não aplicável	184	29.02
Negativo	141	22.24
Ainda não recebeu	21	3.31
Indeterminado	1	0.16

Os dados indicam que aproximadamente 45% dos internados testaram positivo para COVID-19. No entanto, é importante destacar que esse resultado não implica necessariamente que a internação foi motivada pela infecção. Em alguns casos, o entrevistado pode ter testado positivo antes, durante ou após o período de internação. Outro ponto relevante da análise foi a necessidade de entubação para respiração artificial.

entubacao	quantidade	porcentagem
Não	480	75.71
Sim	154	24.29

Entre o número de internados, aproximadamente 24% precisaram desse tipo de suporte. Da mesma forma que foi feito anteriormente, é interessante ver a porcentagem de entrevistados que positivaram para COVID-19 em relação a porcentagem que não apresentou um teste positivo.

entubacao	quantidade	porcentagem
Não	189	65.85
Sim	98	34.15

Pouco mais de 34% dos entrevistados que positivaram para COVID-19 necessitaram de aparelhos de respiração artificial.

entubacao	quantidade	porcentagem
Não	291	83.86
Sim	56	16.14

Em contrapartida, 16,14% dos entrevistados não positivaram para COVID-19 necessitaram de aparelhos de respiração artificial.

A análise demonstra que o uso de respiração artificial foi mais comum entre os positivados para COVID-19. Embora o número de pessoas que testaram positivo tenha sido menor, o número de entubados nesse grupo foi 75% maior do que entre os que não testaram positivo.

diagnostico	total_quantidade	porcentagem
2	285	45.09
7	223	35.28
1	148	23.42
4	113	17.88
3	110	17.41
5	54	8.54
6	32	5.06

- 1 – Diabetes
- 2 – Hipertensão
- 3 – Asmas, bronquite, efisema, doenças respiratórias crônicas ou doenças do pulmão
- 4 – Doenças do coração
- 5 – Depressão
- 6 – Câncer
- 7 - Nenhum

Quanto aos diagnósticos mais comuns entre os internados, verificou-se que a hipertensão prevaleceu entre os entrevistados que testaram positivo para COVID-19, estando presente em 45% dos casos. Além disso, 35% dos casos de hipertensão estavam associados a outros diagnósticos, sendo o mais comum a diabetes, que apareceu em 20% das respostas (56 afirmações). Entre os não positivados, as proporções de diagnósticos também seguiram um padrão similar.

diagnostico	total_quantidade	porcentagem
2	151	43.64
7	114	32.95
1	79	22.83
3	75	21.68
4	72	20.81
5	38	10.98
6	27	7.8

Além de identificar o diagnóstico mais comum, foi analisada a taxa de internação associada a cada diagnóstico para os pacientes positivados.

Resultados para diagnóstico: Diabetes

internacao	diagnostico_1	porcentagem
Não	37	5.24
Não aplicável	659	93.34
Sim	10	1.42

Resultados para diagnóstico: Hipertensão

internacao	diagnostico_2	porcentagem
Não	150	4.97
Não aplicável	2810	93.05
Sim	59	1.95
ignorado	1	0.03

Resultados para diagnóstico: Asma/bronquite/enfisema/doenças respiratórias crônicas ou doença de pulmão

internacao	diagnostico_3	porcentagem
Não	66	5.62
Não aplicável	1097	93.36
Não foi atendido	1	0.09
Sim	11	0.94

Resultados para diagnóstico: Doenças do coração (infarto, angina, insuficiência cardíaca, arritmia)		
internacao	diagnostico_4	porcentagem
Não	14	5.17
Não aplicável	252	92.99
Sim	5	1.85
Resultados para diagnóstico: Depressão		
internacao	diagnostico_5	porcentagem
Não	42	7.65
Não aplicável	503	91.62
Sim	4	0.73
Resultados para diagnóstico: Câncer		
internacao	diagnostico_6	porcentagem
Não	6	4.32
Não aplicável	130	93.53
Não foi atendido	1	0.72
Sim	2	1.44

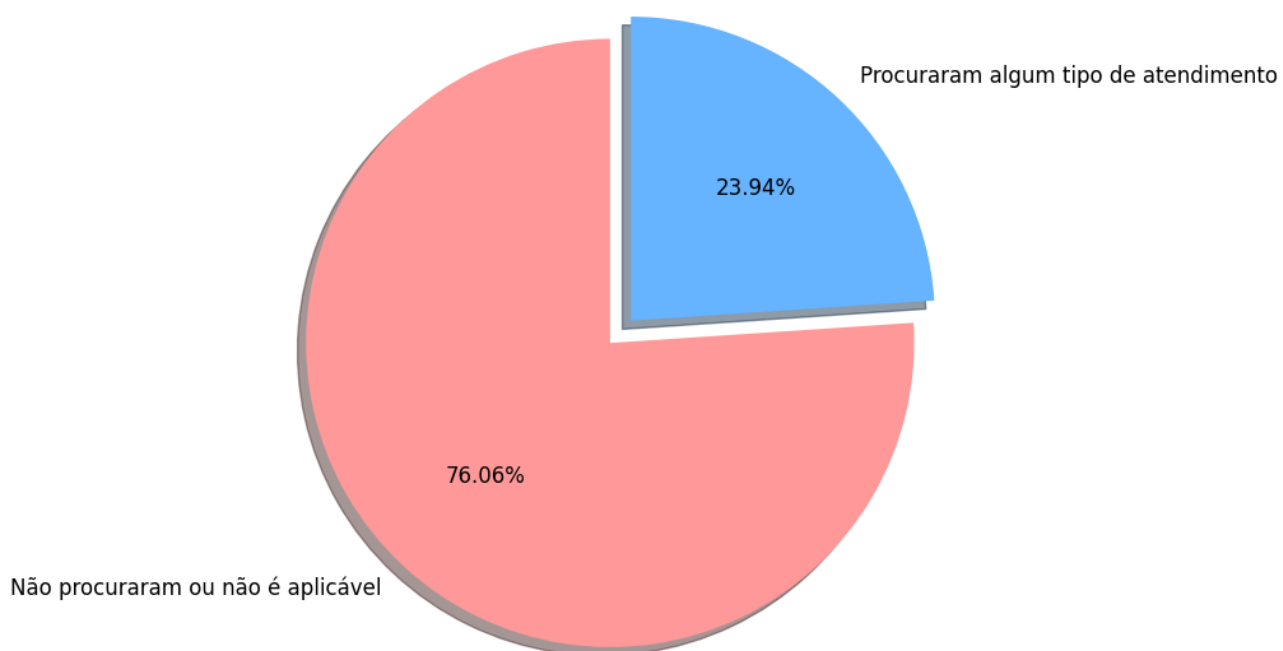
Entre todos os diagnósticos presentes no questionário, a hipertensão apresentou a maior necessidade de internação, com uma taxa de 1,95%, seguida pelas doenças do coração (1,85%) e o câncer (1,44%).

4 Possíveis Ações

A partir das análises realizadas, é possível identificar ações estratégicas para mitigar a disseminação do COVID-19 em um possível novo surto. Na seção 3.1, observou-se que mais de 95% dos entrevistados que não adotaram qualquer tipo de restrição e não realizaram testes de COVID relataram não ter apresentado sintomas. No entanto, o ponto crítico reside nos quase 5% que, apesar de apresentarem sintomas, não tomaram nenhuma medida preventiva e tampouco realizaram testes. Esse grupo representa um potencial risco de transmissão do vírus.

Além disso, aproximadamente 24% dessas pessoas procuraram atendimento médico, conforme evidenciado nas análises visuais. Isso indica que uma parcela significativa dos possíveis transmissores não foi identificada e isolada.

Relação de entrevistados que não fizeram restrição e teste de COVID que procuraram atendimento



Com base nessas informações, uma ação eficaz seria ampliar o acesso a testes de COVID, especialmente para os entrevistados que apresentaram sintomas e buscaram alguma forma de atendimento. Essa medida ajudaria a identificar rapidamente novos casos e prevenir a propagação do vírus, diminuindo o risco de contágio na comunidade.

Outra possível ação surge a partir da análise 3.3, que avaliou a quantidade de internações por diagnóstico. Essa análise permite um acompanhamento mais rigoroso de grupos específicos que apresentam maior probabilidade de necessitar internação em caso de contaminação por COVID-19, como, por exemplo, os pacientes com hipertensão ou doenças do coração.

Esses grupos de risco, identificados pela taxa de internação, podem ser priorizados em estratégias de monitoramento e cuidados preventivos. Acompanhamentos regulares e medidas mais intensivas de proteção e tratamento poderiam reduzir significativamente a necessidade de hospitalizações e complicações graves entre essas pessoas, aliviando a pressão sobre os serviços de saúde.