Jonathan Abner Jerônimo de Freitas jonathanabner2015@gmail.com
Osvaldo Caio Oliveira dos Santos osvaldocaio@hotmail.com
Ranielli Santos raniellidos@hotmail.com
Thalita Mendes Maina Begliomini thalitamaina@gmail.com

_ROTEIRO

DEFINIÇÕES

PROBLEMA

CORONAVÍRUS

01

_ DEFINIÇÃO DO PROBLEMA

Imagine agora que você foi contratado(a) como Expert em Data Analytics por um grande hospital para entender como foi o comportamento da população na época da pandemia da COVID-19 e quais indicadores seriam importantes para o planejamento, caso haja um novo surto da doença.

Apesar de ser contratado(a) agora, a sua área observou que a utilização do estudo do PNAD-COVID 19 do IBGE seria uma ótima base para termos boas respostas ao problema proposto, pois são dados confiáveis.

Porém, não será necessário utilizar todas as perguntas realizadas na pesquisa para enxergar todas as oportunidades ali postas. É sempre bom ressaltar que há dados triviais que precisam estar no projeto, pois auxiliam muito na análise dos dados:

/ Características clínicas dos sintomas;

/ Características da população;

/ Características econômicas da sociedade.

_ DEFINIÇÃO DO PROBLEMA

PNAD-COVID-19 do IBGE O Head de Dados pediu para que você entrasse na base de dados do PNAD-COVID-19 do IBGE e organizasse esta base para análise, utilizando Banco de Dados em Nuvem e trazendo as seguintes características:

a. Utilização de no máximo 20 questionamentos realizados na pesquisa;

b. Utilizar 3 meses para construção da solução;

c. Caracterização dos sintomas clínicos da população;

d. Comportamento da população na época da COVID-19;

e. Características econômicas da Sociedade;

Seu objetivo será trazer uma breve análise dessas informações, como foi a organização do banco, as perguntas selecionadas para a resposta do problema e quais seriam as principais ações que o hospital deverá tomar em caso de um novo surto de COVID-19.

Dica: Leiam com atenção a base de dados

_ SOBRE A PNAD COVID19

O que é?

Objetiva estimar o número de pessoas com sintomas referidos associados à síndrome gripal e monitorar os impactos da pandemia da COVID-19 no mercado de trabalho brasileiro.

A coleta da Pesquisa Nacional por Amostra de Domicílios - PNAD COVID19 teve início em 4 de maio de 2020, com entrevistas realizadas por telefone em, aproximadamente, 48 mil domicílios por semana, totalizando cerca de 193 mil domicílios por mês, em todo o Território Nacional. A amostra é fixa, ou seja, os domicílios entrevistados no primeiro mês de coleta de dados permanecerão na amostra nos meses subsequentes, até o fim da pesquisa.

_ CORONAVÍRUS

A Covid-19 é uma infecção respiratória aguda causada pelo coronavírus SARS-CoV-2, potencialmente grave, de elevada transmissibilidade e de distribuição global. O SARS-CoV-2 é um betacoronavírus descoberto em amostras de lavado broncoalveolar obtidas de pacientes com pneumonia de causa desconhecida na cidade de Wuhan, província de Hubei, China, em dezembro de 2019. Pertence ao subgênero Sarbecovírus da família Coronaviridae e é o sétimo coronavírus conhecido a infectar seres humanos.

De acordo com as evidências mais atuais, o SARS-CoV-2, da mesma forma que outros vírus respiratórios, é transmitido principalmente por três modos: contato, gotículas ou por aerossol.

A principal medida de prevenção contra formas graves da covid-19 é a vacina.

A campanha de vacinação contra a covid-19 foi iniciada em janeiro de 2021!

SOBRE O PROJETO

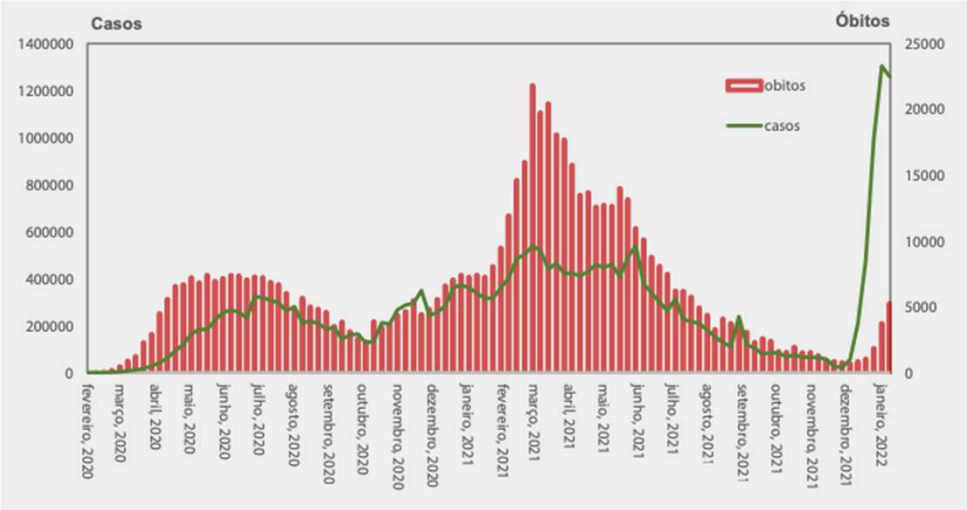
02

_ SOBRE O PROJETO

Banco de dados - O banco de dados foi disponibilizado no site para pesquisa no PNAD-COVID-19 do IBGE.

Foi inserido 3 tabelas de dados referentes aos meses de julho, agosto e setembro de 2020.

Utilizado estes meses, devido ao maior período de alta e posterior queda no ano de 2020. Todos estes dados foram incluídos e estruturados no Google Cloud.



Estrutura - O banco de dados SQL 'pos_tech3_caio' criado foi composto por 3 tabelas PNAD 6, PNAD 7 e PNAD 8 com uma tabela que armazena todos os dados da pesquisa.

A pesquisa é extensa contendo 1.151.956 linhas e 20 colunas, já nosso campo de pesquisa se limitou a trabalhar com as variáveis mostradas no próximo slide: Dicionário.

_ DICIONÁRIO

Código das variáveis escolhidas para análise e sua descrição

Código da variável	nº	Descrição
Ano		Ano de referência
UF		Unidade da Federação
V1013		Mês da pesquisa
V1022		Situação do domicílio
A002	A2	Idade do morador
A003	A3	Sexo
A004	A4	Cor ou raça
A005	A5	Escolaridade
B002	B2	Por causa disso, foi a algum estabelecimento de saúde?
B0031	B3	Providência tomada para recuperar dos sintomas foi ficar em casa
B0032	B3	Providência tomada para recuperar dos sintomas foi ligar para algum profissional de saúde
B0042	B4	Local que buscou atendimento foi pronto socorro do SUS/UPA
B0043	B4	Local que buscou atendimento foi hospital do SUS
B005	B5	Ao procurar o hospital, teve que ficar internado por um dia ou mais
B007	B7	Tem algum plano de saúde médico, seja particular, de empresa ou de órgão público
C003	C3	Qual o principal motivo deste afastamento temporário?
C013	C13	Na semana passada, o(a) Sr(a) estava em trabalho remoto (home office ou teletrabalho)?
C016	C16	Qual o principal motivo de não ter procurado trabalho na semana passada?
D0031	D1	Rendimentos de Programa Bolsa Família
D0061	D1	Seguro desemprego
F001	F1	Este domicílio é:

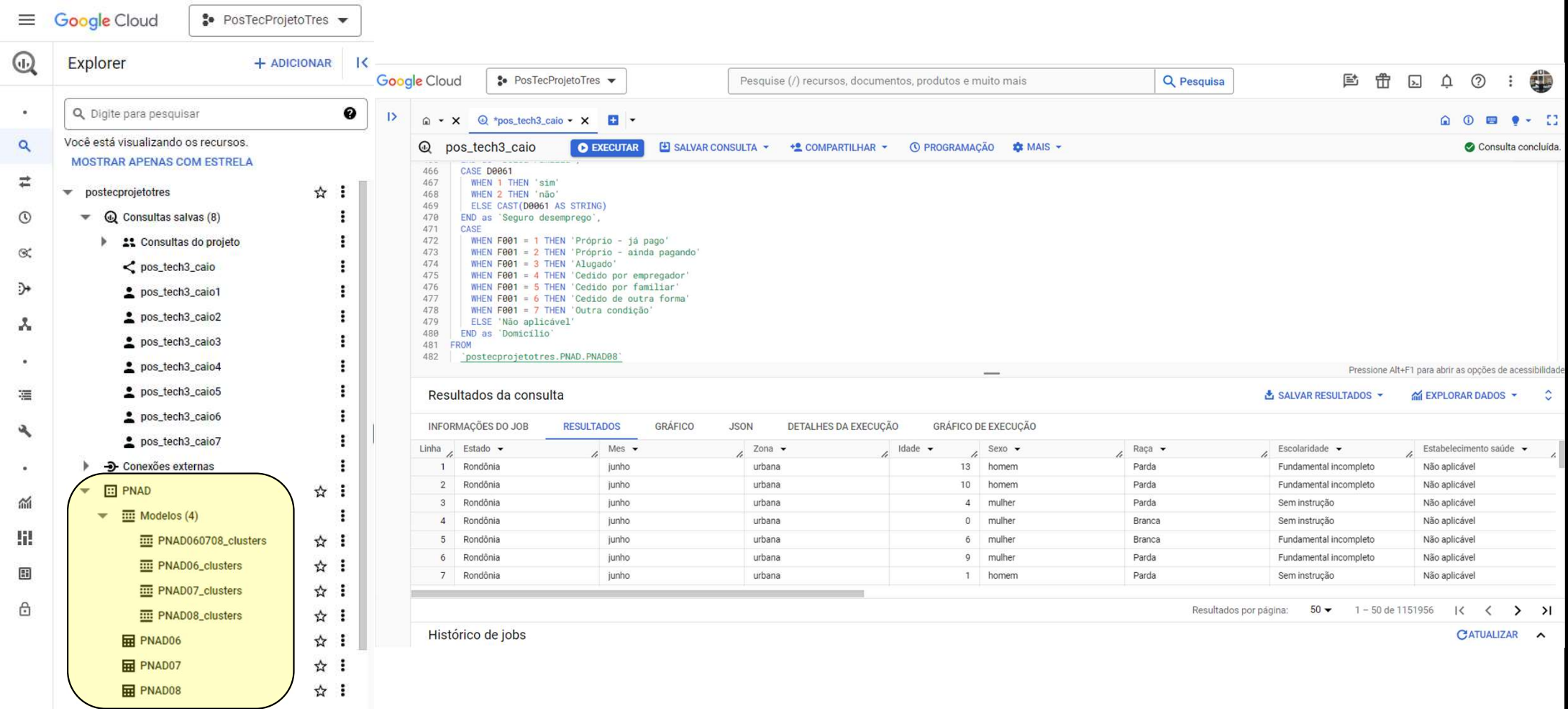
Respostas: As opções de respostas e regras são mostradas no link: PNAD COVID19 – Questionário

No trabalho as opções apontadas como 'Não Aplicável' se dá ao fato de serem ignoradas no momento da pesquisa, seja devido as regras das pesquisas ou usuário.

_GOOGLE CLOUD

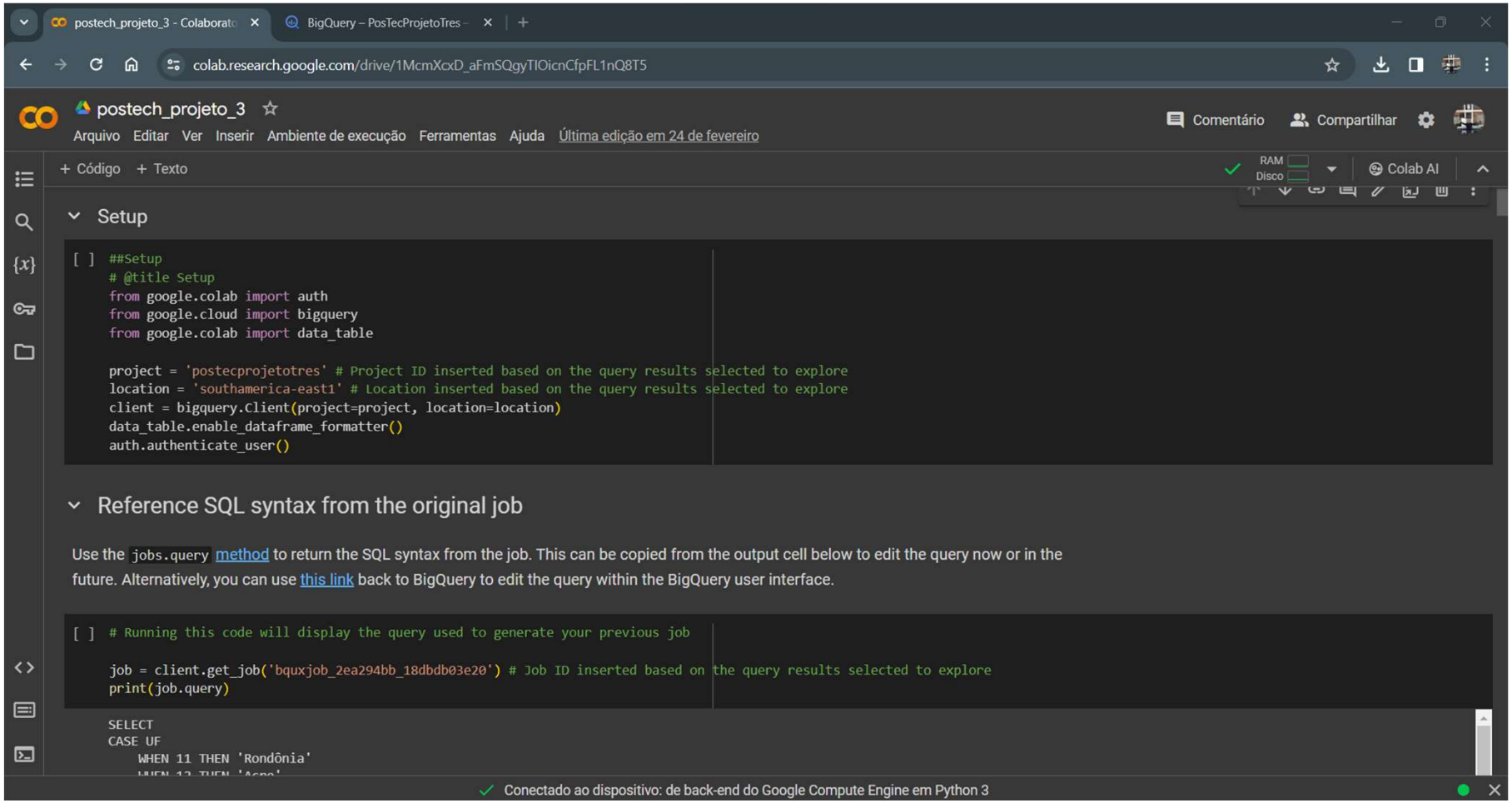
Carregado dados para o Google Cloud conforme podemos verificar no destaque amarelo.

Alterado por código as variáveis da tabela para descrição conforme imagem abaixo.



_GOOGLE COLAB

Após código SQL pronto e tabela OK no Google Cloud, foi transferido para o Google Colab, por meio de uma conexão python. Obs.: essa transferência é automática.



_GOOGLE COLAB

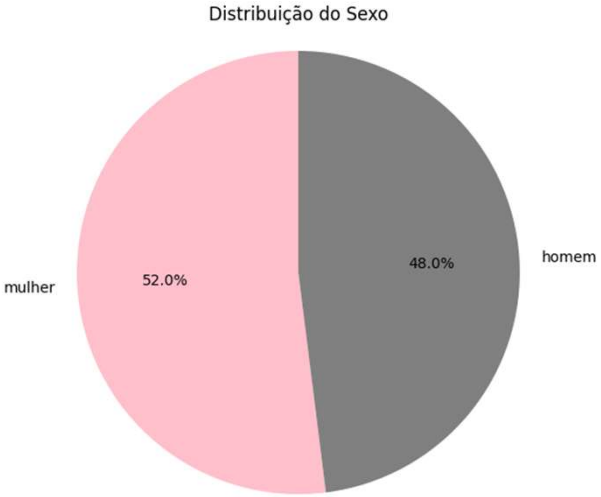
Imagem da tabela de dados no Google Colab e a quantidade de dados.

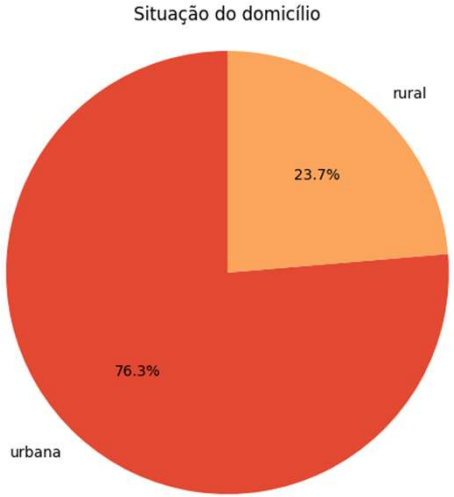
	Estado	Mes	Zona	Idade	Sexo	Raça	Escolaridade	Estabelecimento saúde	Providência - casa	Providência - profissional	Atendimento PS	Atendimento hospital	Internação	Plano de saúde	Afastamento	Trabalho remoto	PQ não procurou trabalho	Bolsa Familia	Seguro desemprego	Domicílio
0	Rondônia	junho	urbana	3	homem	Parda	Sem instrução	Não	Sim	Não	Não aplicável	Não aplicável	Não aplicável	Não	Não aplicável	Não aplicável	Não aplicável	não	não	Próprio - ainda pagando
1	Rondônia	junho	urbana	7	homem	Branca	Fundamental incompleto	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Sim	Não aplicável	Não aplicável	Não aplicável	não	não	Alugado
2	Rondônia	junho	urbana	8	homem	Parda	Fundamental incompleto	Não	Sim	Não	Não aplicável	Não aplicável	Não aplicável	Não	Não aplicável	Não aplicável	Não aplicável	não	não	Cedido por familiar
3	Rondônia	junho	urbana	4	mulher	Parda	Sem instrução	Não	Sim	Não	Não aplicável	Não aplicável	Não aplicável	Não	Não aplicável	Não aplicável	Não aplicável	sim	não	Alugado
4	Rondônia	junho	urbana	2	mulher	Parda	Sem instrução	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não	Não aplicável	Não aplicável	Não aplicável	sim	não	Cedido por familiar
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1151951	São Paulo	agosto	urbana	29	homem	Parda	Fundamental completa	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não	Estava em quarentena, isolamento, distanciamen...	Não aplicável	Não aplicável	não	não	Alugado
1151952	Paraná	agosto	urbana	39	mulher	Branca	Médio completo	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Sim	Estava em quarentena, isolamento, distanciamen...	Não aplicável	Não aplicável	não	sim	Próprio - já pago
1151953	Paraná	agosto	rural	51	mulher	Parda	Fundamental incompleto	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Sim	Licença remunerada por motivo de saúde ou acid...	Não aplicável	Por problemas de saúde ou gravidez	não	não	Cedido por empregador
1151954	Santa Catarina	agosto	urbana	47	mulher	Branca	Médio completo	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não	Estava em quarentena, isolamento, distanciamen...	Não aplicável	Não quer trabalhar ou é aposentado	não	não	Próprio - já pago
1151955	Distrito Federal	agosto	urbana	20	homem	Parda	Superior incompleto	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não aplicável	Não	Estava em quarentena, isolamento, distanciamen...	Não aplicável	Devido à pandemia (isolamento, quarentena ou d...	não	não	Próprio - já pago

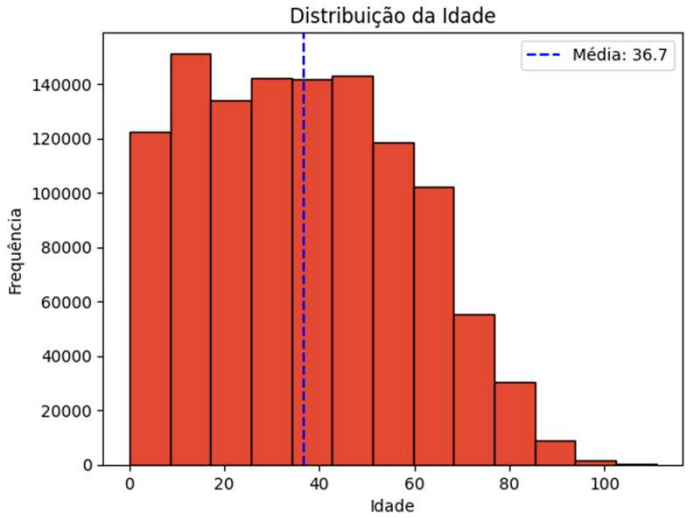
1151956 rows × 20 columns

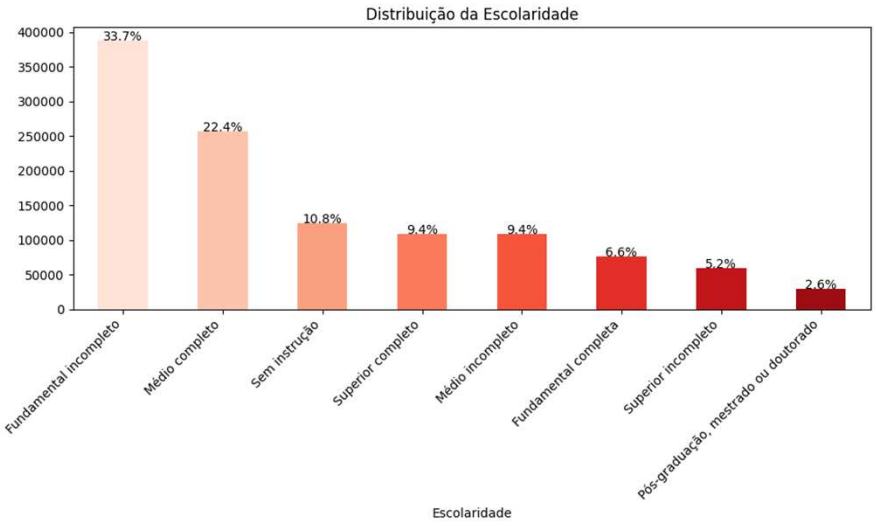
Criado todos os gráficos a seguir em python no Google Colab.

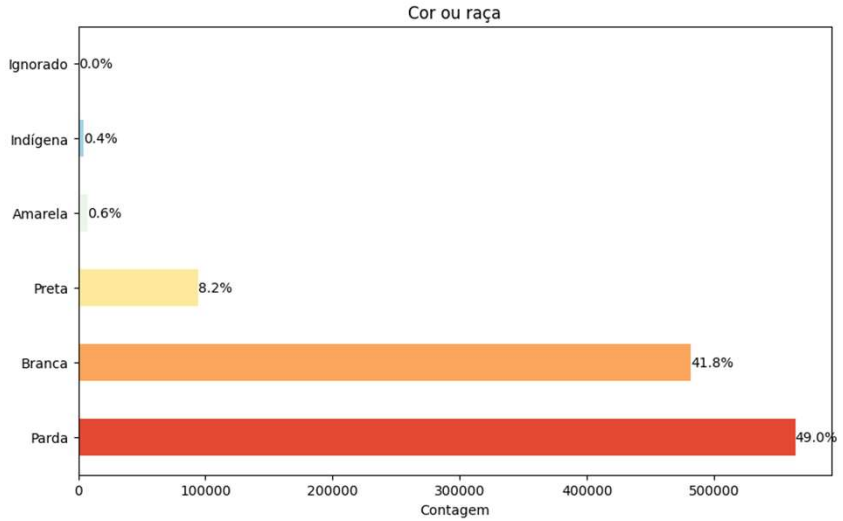
CARACTERÍSTICAS GERAIS



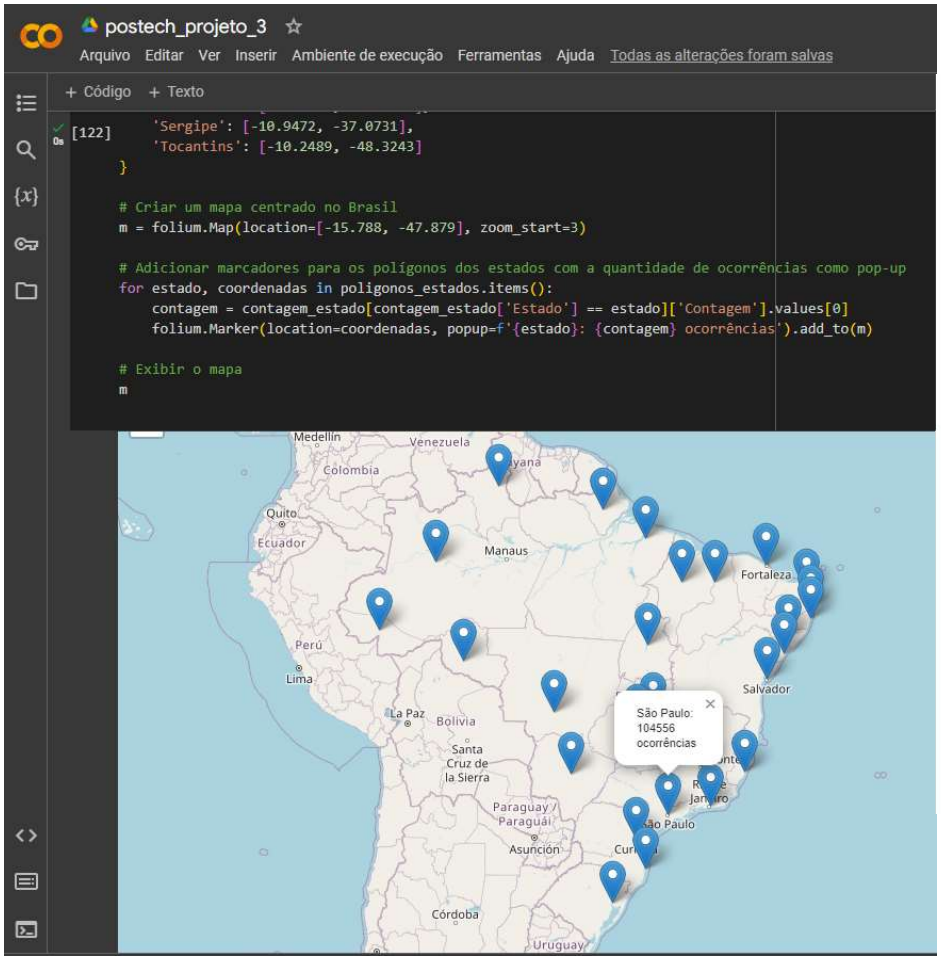


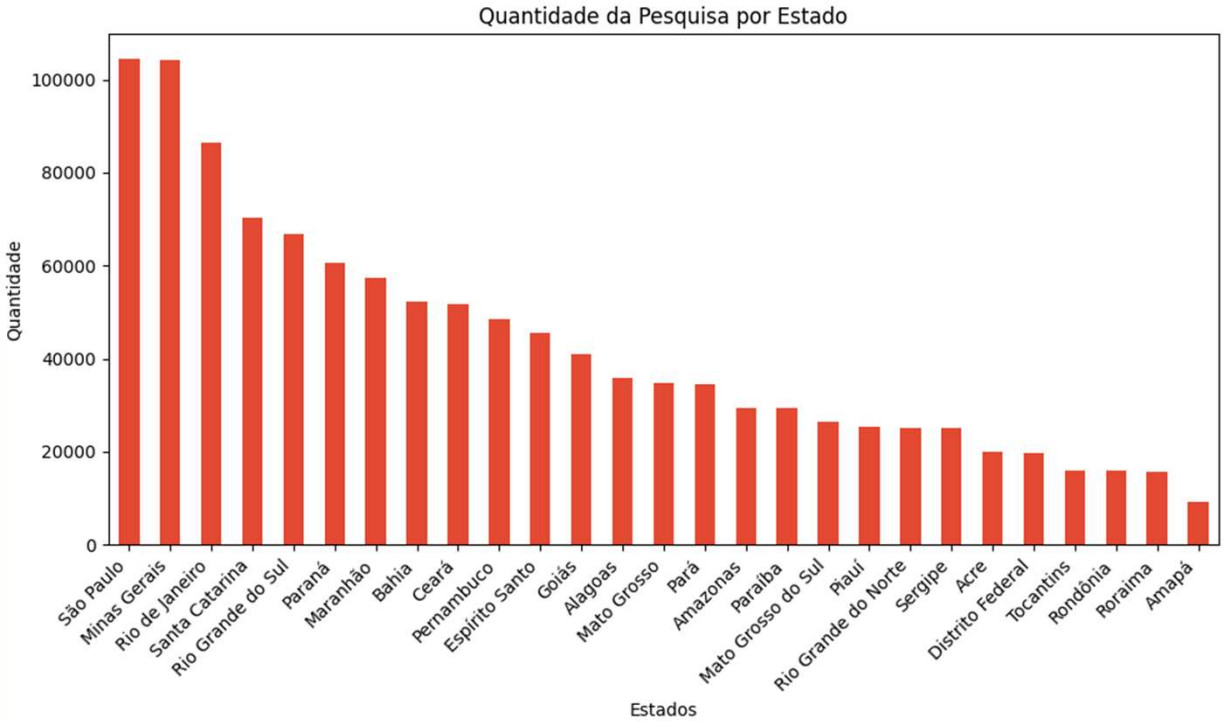




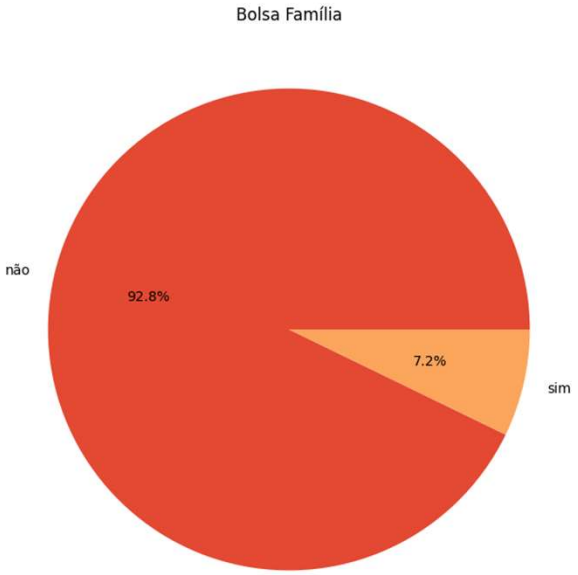


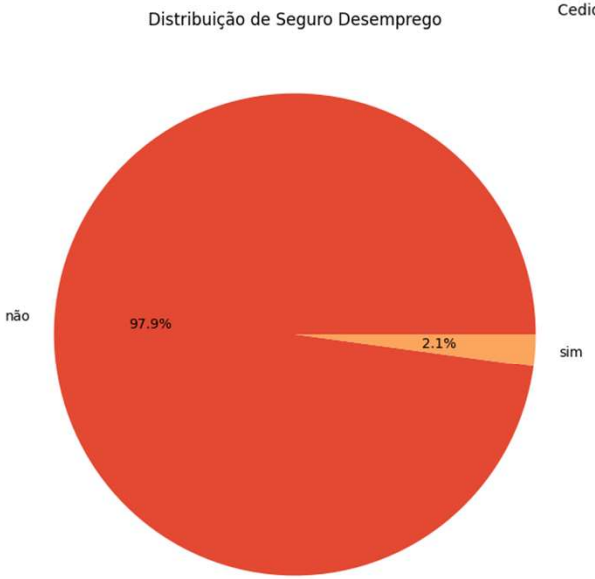
_ CARACTERÍSTICAS DE LOCALIZAÇÃO

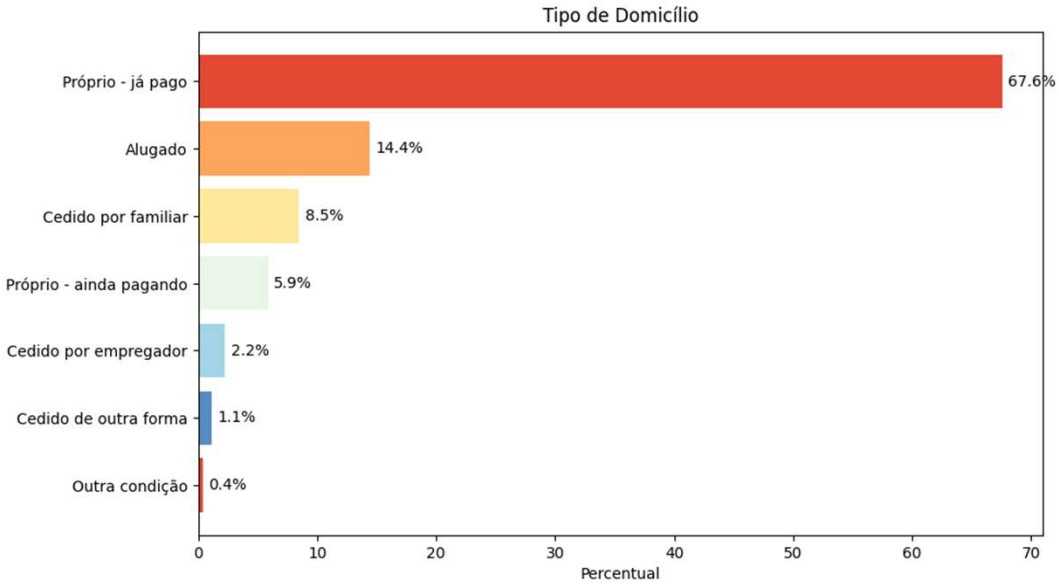




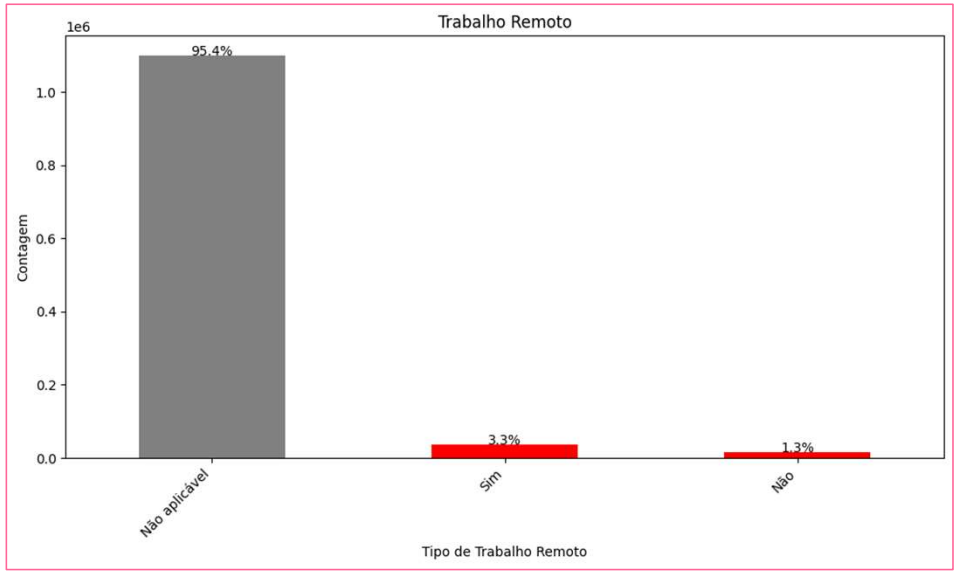
_ CARACTERÍSTICAS ECONÔMICAS

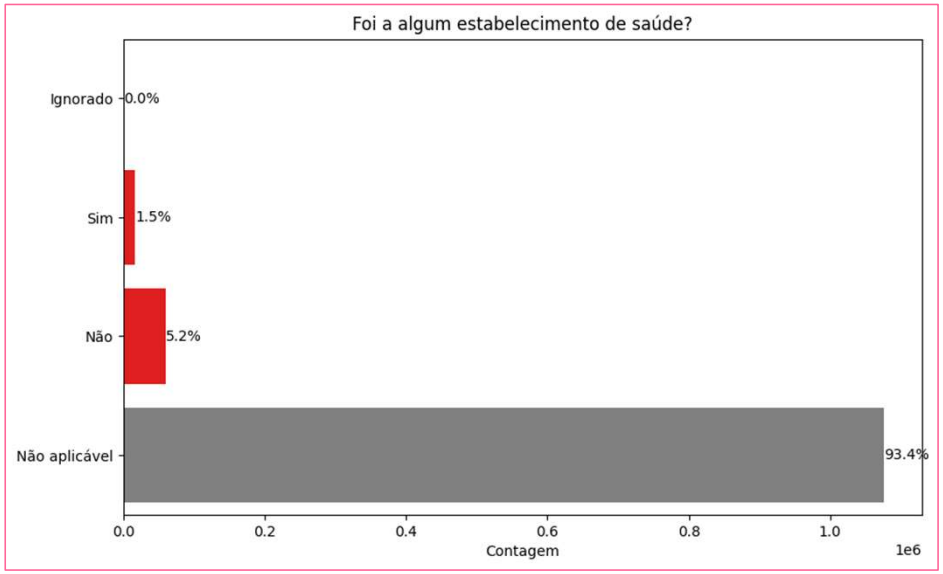


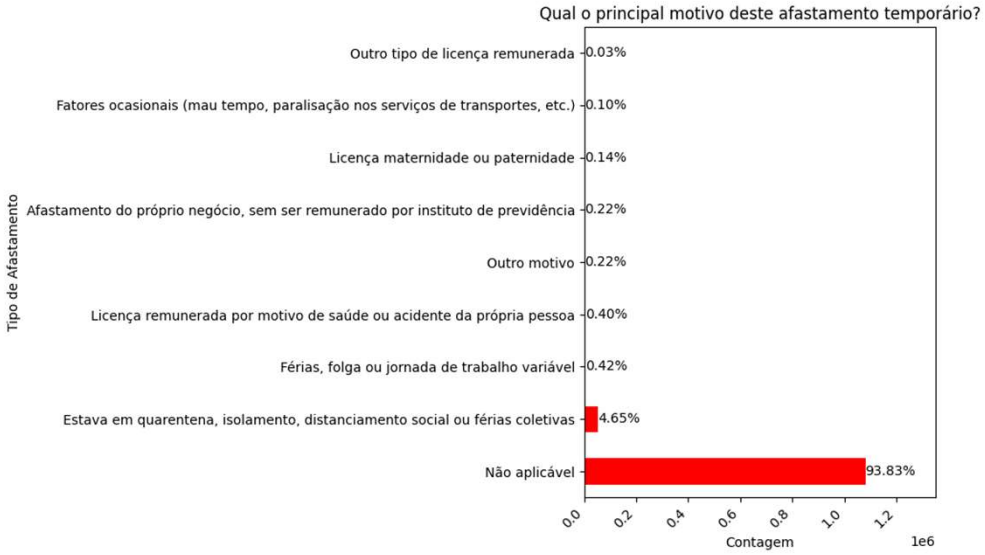




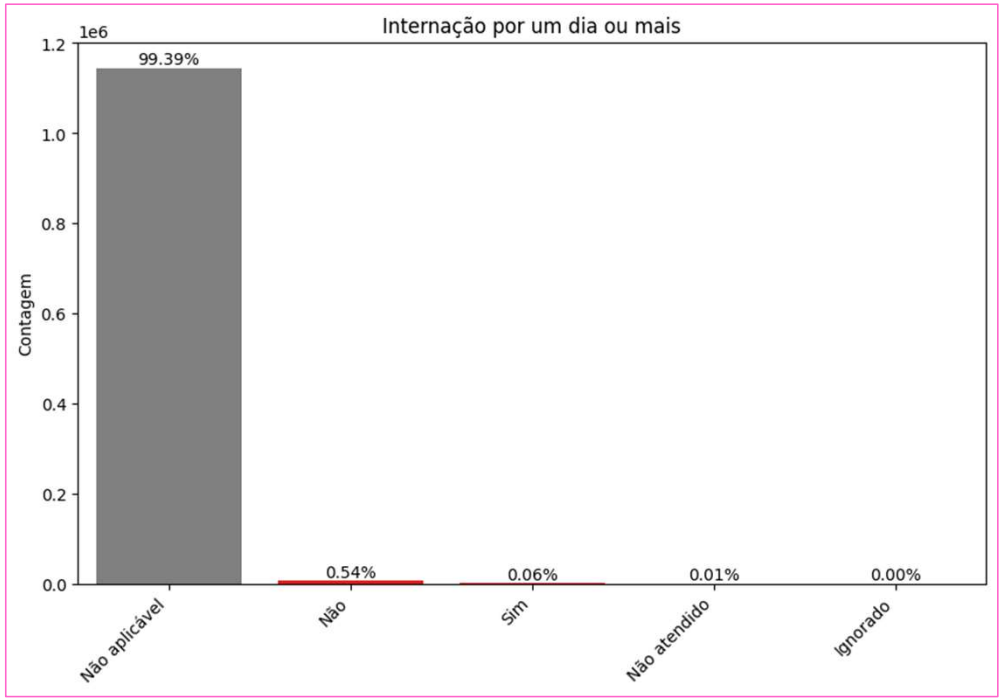
CARACTERÍSTICAS GERAIS

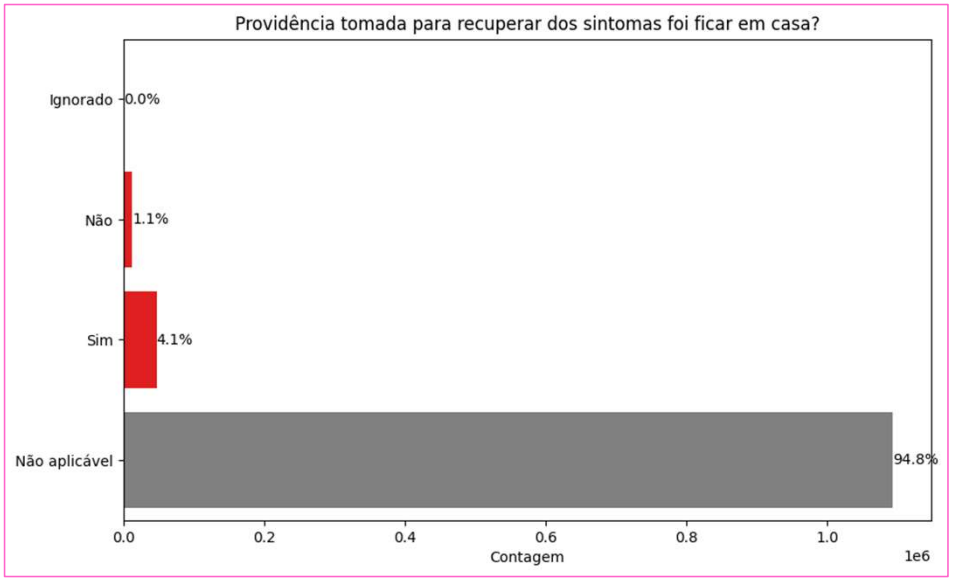


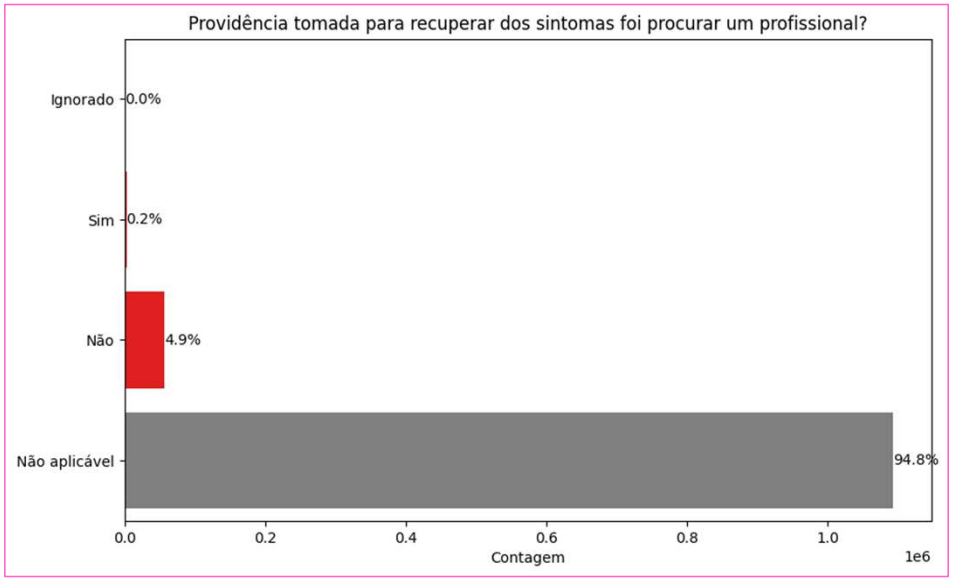




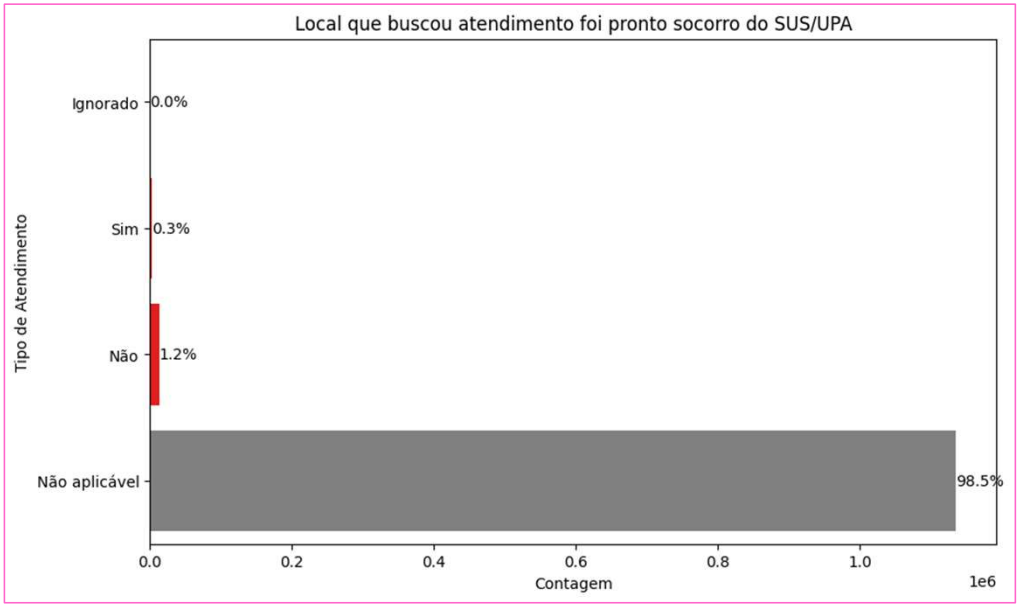
CARACTERÍSTICAS CLÍNICAS

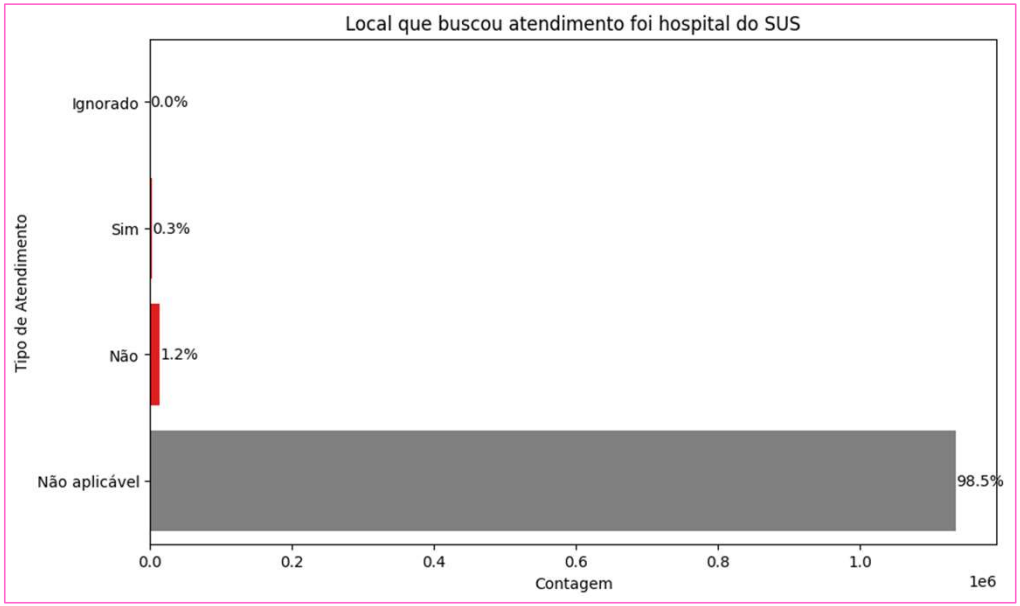






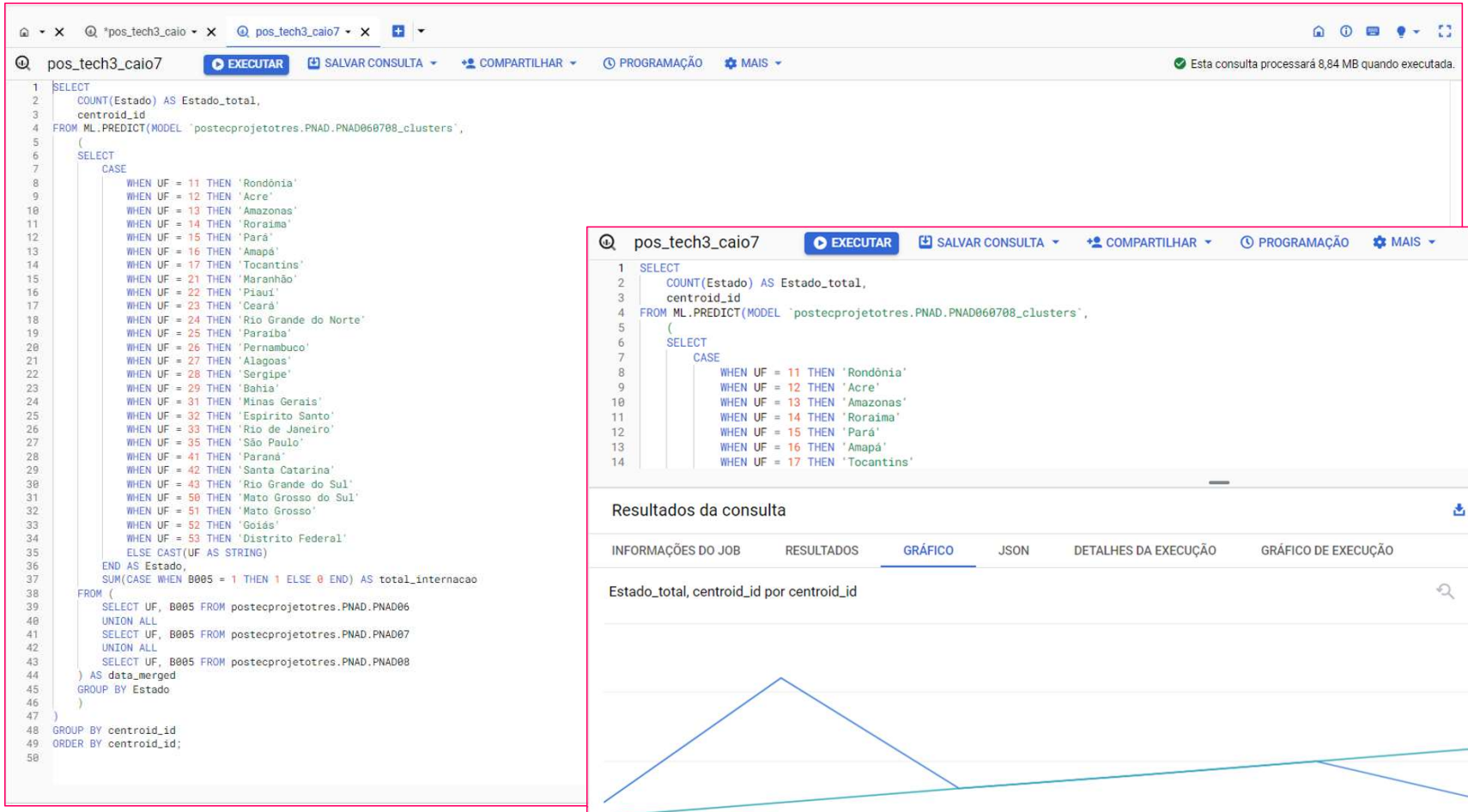
_ CARACTERÍSTICAS CLÍNICAS





_ CLUSTERING-ML

Mostrado agrupamento dos estados e o total de internação mostrando a similaridade dos dados entre os 3 meses.



CONSIDERAÇÕES FINAIS

04

_ CONSIDERAÇÕES FINAIS

Após uma análise detalhada do comportamento da população na época de pandemia da COVID-19, recomendamos os seguintes indicadores para uma melhor gestão de um grande hospital para se planejar de forma mais adequada em novos surtos:

Alguns dos atributos de maior peso para monitoração do Público em atendimento no hospital:

idade, sexo, etnia, status socioeconômico, sintomas e comorbidades
% atendimento diário no ambiente/ moradia urbana
% público com menor grau de escolaridade sendo sem instrução, médio completo ou fundamental incompleto
% público sem plano de saúde

_ CONSIDERAÇÕES FINAIS

Além destes indicadores combinados com outros dados para atuar de forma preventiva:

1. Uso de sistema de saúde. Observar as tendências como visitas no pronto socorro e internações devido ao vírus pode apoiar no entendimento da gravidade da doença;

2- Dados de mobilidade para entender padrões de deslocamento da população e permitir que o hospital aloque recursos;

3- Monitoramento de mídias sociais na adesão a medidas preventivas para apoiar em ideias de conscientização e ajustes na comunicação;

4- Monitoramento de indicadores de saúde pública como taxa de positividade de testes.

De maneira simplificada, um modelo para apoiar o hospital deverá se concentrar uma grande quantidade de dados que, cruzados, indicam padrões e tendências futuras sobre os mais diferentes problemas, como a disseminação de doenças.

O modelo proposto pode contribuir para a diminuição da subjetividade na tomada de decisão, para que o profissional de saúde possa tomar suas decisões baseadas em dados de forma mais assertiva.

REFERÊNCIAS

05

REFERÊNCIAS

- https://covid19.ibge.gov.br/pnad-covid/
- https://www.fiocruzbrasilia.fiocruz.br/covid-19-balanco-de-dois-anos-da-pandemia-aponta-vacinacao-como-prioridade/
- https://www.ibge.gov.br/estatisticas/investigacoes-experimentais/estatisticas-experimentais/27946-divulgacao-semanal-pnadcovid1?t=o-que-e&utm_source=covid19&utm_medium=hotsite&utm_campaign=covid_19
- https://www.gov.br/saude/pt-br/assuntos/coronavirus
- https://www.ibge.gov.br/estatisticas/investigacoes-experimentais/estatisticas-experimentais/27946-divulgacao-semanal-pnadcovid1?t=downloads&utm_source=covid19&utm_medium=hotsite&utm_campaign=covid_19