



TECH CHALLENGE

FASE 03

6DTAT - Grupo 44



TÓPICOS

1. Objetivos
2. Base de Dados
3. Tratamento no BigQuery
4. Análise de Dados
 - I. Caracterização dos sintomas clínicos;
 - II. Comportamento da população na época da COVID;
 - III. Características econômicas da Sociedade
5. Conclusão



OBJETIVOS

Objetivo principal

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.

Objetivos secundários

Objetivo 1	Utilização de no máximo 20 questionamentos realizados na pesquisa
Objetivo 2	Utilizar 3 meses para construção da solução
Objetivo 3	Caracterização dos sintomas clínicos da população
Objetivo 4	Comportamento da população na época da COVID-19
Objetivo 5	Características econômicas da Sociedade;

BASE DE DADOS

Informações

As informações iniciais estão disponíveis no PNAD-COVID-19, publicado pelo IBGE (<https://covid19.ibge.gov.br/pnad-covid/>).

Foram coletadas informações de cerca de 193 mil pessoas no período entre Maio/20 e Novembro/20, através de um questionário padronizado que será detalhado mais a frente.

Essa mesma base de dados consta na biblioteca do BigQuery, através do ID [basedosdados.br_ibge_pnad_covid.microdados](#), disponibilizada pela equipe do site [basedosdados.org](#), e foi utilizada para o desenvolvimento deste trabalho.

BASE DE DADOS

Questionário

As informações foram obtidas pelo IBGE através de um questionário com diversas perguntas, realizadas semanalmente, e disponibilizado no portal do PNAD-19. Ao longo do período entre Maio/20 e Novembro/20, foram três questionários diferentes.

Para que tenhamos uniformidade nas análises, os dados que serão apresentados correspondem à 03 meses em que foram aplicados os mesmos questionários, portanto nossa base abrangerá as respostas entre **Julho/20 e Setembro/20**.

BASE DE DADOS

Questionário

PNAD COVID19

(passe ao módulo B)

Módulo B – COVID19 – Todos os moradores

Entrevistador leia: Vamos fazer algumas perguntas sobre o novo coronavírus. Primeiro os sintomas. Vou dizer os sintomas e você me responde “sim” ou “não”.

B1. Na semana passada, o(a) Sr(a) teve:

a. Febre	1. Sim	2. Não	3. Não sabe (B0011)
b. Tosse	1. Sim	2. Não	3. Não sabe (B0012)
c. Dor de garganta	1. Sim	2. Não	3. Não sabe (B0013)
d. Dificuldade para respirar	1. Sim	2. Não	3. Não sabe (B0014)
e. Dor de cabeça	1. Sim	2. Não	3. Não sabe (B0015)
f. Dor no peito	1. Sim	2. Não	3. Não sabe (B0016)
g. Náusea/enjoo	1. Sim	2. Não	3. Não sabe (B0017)
h. Nariz entupido ou escorrendo	1. Sim	2. Não	3. Não sabe (B0018)
i. Fadiga/cansaço	1. Sim	2. Não	3. Não sabe (B0019)
j. Dor nos olhos	1. Sim	2. Não	3. Não sabe (B00110)
k. Perda de cheiro ou de sabor	1. Sim	2. Não	3. Não sabe (B00111)
l. Dor muscular/ dor no corpo	1. Sim	2. Não	3. Não sabe (B00112)
m. Diarreia	1. Sim	2. Não	3. Não sabe (B00113)

(Se a ou b ou c ou d ou e ou f ou g ou h ou i ou j ou k ou l ou m = 1, siga ao B2. Caso contrário, passe B7.)

Entrevistador leia: Agora vou perguntar sobre o seu comportamento em relação a esses sintomas.

B2. Por causa disso, o(a) Sr(a) foi a algum estabelecimento de saúde (na semana passada)? (B002)

1. Sim 2. Não

(Se B2 = 1, passe ao B4)

(Se B2 = 2, siga B3)

(Se ignorado, passe ao B7)

Ao todo, foram realizadas 53 perguntas, conforme modelo ao lado.

Para a análise, foram selecionadas 20 dessas perguntas para maior objetividade e eficiência nas consultas.

BASE DE DADOS

Dicionário de Dados

Dicionário das variáveis da PNAD COVID					
Tamanho	Código da variável	Quesito		Categorias	
		nº	Descrição	Tipo	Descrição
Parte B - COVID19 - Todos os moradores					
1	B0011	B1	Na semana passada teve febre?	1	Sim
				2	Não
				3	Não sabe
				9	Ignorado
1	B0012	B1	Na semana passada teve tosse?	1	Sim
				2	Não
				3	Não sabe
				9	Ignorado
1	B0013	B1	Na semana passada teve dor de garganta?	1	Sim
				2	Não
				3	Não sabe
				9	Ignorado
1	B0014	B1	Na semana passada teve dificuldade para respirar?	1	Sim
				2	Não
				3	Não sabe
				9	Ignorado
1	B0015	B1	Na semana passada teve dor de cabeça?	1	Sim
				2	Não
				3	Não sabe
				9	Ignorado
1	B0016	B1	Na semana passada teve dor no peito?	1	Sim
				2	Não
				3	Não sabe
				9	Ignorado
1	B0017	B1	Na semana passada teve náusea?	1	Sim
				2	Não
				3	Não sabe
				9	Ignorado
1	B0018	B1	Na semana passada teve nariz entupido ou escorrendo?	1	Sim
				2	Não
				3	Não sabe
				9	Ignorado

As colunas na base possuem um código alfanumérico para identificação, assim como a maior parte das respostas também são padronizadas numericamente.

Para melhor entendimento da base, foi disponibilizado no portal do PNAD COVID-19 um dicionário de dados, utilizado no desenvolvimento para melhor compreensão.

BASE DE DADOS

Questões Selecionadas

A2. Idade	B4. Local que buscou atendimento	B9D. Resultado do Teste de furo no dedo	C7b. Tem carteira de trabalho assinada?
A3. Sexo	B5. Ocorreu internação?	B9F. Resultado do Teste de veia no braço	C10. Quanto recebe normalmente?
A5. Escolaridade	B6. Foi entubado e colocado em respiração artificial?	B10. Comorbidades	C11a. Quanto recebeu no mês vigente?
B1. Sintomas	B7. Possui Plano de Saúde?	B11. Medidas de restrição	E1. Solicitou empréstimo?
B2. Comparecimento à estabelecimento de Saúde	B9B. Resultado do Teste de Swab	C1. Trabalhou ou realizou algum bico?	F2a. No seu domicílio, há quais itens básicos de limpeza e proteção?

TRATAMENTO NO BIGQUERY

Foi utilizado o código abaixo para realizar o pré-tratamento da base e filtragem dos dados de interesse, que posteriormente foram utilizados em um Notebook via Google Colaboratory.

```
Consulta_Filtrada EXECUTAR SALVAR CONSULTA FAZER O DOWNLOAD

1 SELECT ano,
2       mes,
3       semana,
4       (sigla_uf) AS uf,
5       CAST(A002 AS INT64) AS idade,
6       (A003) AS sexo,
7       CAST(A005 AS INT64) AS escolaridade,
8       CAST(B0011 AS INT64) AS teve_febre,
9       CAST(B0012 AS INT64) AS teve_tosse,
10      CAST(B0013 AS INT64) AS teve_dor_garganta,
11      CAST(B0014 AS INT64) AS teve_dificuldade_respirar,
12      CAST(B0015 AS INT64) AS teve_dor_cabeca,
13      CAST(B0016 AS INT64) AS teve_dor_peito,
14      CAST(B0017 AS INT64) AS teve_nausea,
15      CAST(B0018 AS INT64) AS teve_nariz_entupido_escorendo,
16      CAST(B0019 AS INT64) AS teve_fadiga,
17      CAST(B00110 AS INT64) AS teve_dor_olhos,
18      CAST(B00111 AS INT64) AS teve_perda_cheiro_sabor,
19      CAST(B00112 AS INT64) AS teve_dor_muscular,
20      CAST(B00113 AS INT64) AS teve_diarreia,
21      CAST(B002 AS INT64) AS comparecimento_estabelecimento_saude,
22      CAST(B0041 AS INT64) AS postodesaude_unidadebasica_saudeafamilia,
23      CAST(B0042 AS INT64) AS ps_sus_upa,
24      CAST(B0043 AS INT64) AS hospital_sus,
25      CAST(B0044 AS INT64) AS ambulatorio_consultorio_privado_ou_forcasarmadas,
26      CAST(B0045 AS INT64) AS psprivado_forcasarmadas,
27      CAST(B0046 AS INT64) AS hospitalprivado_forcasarmadas,
28      CAST(B005 AS INT64) AS ficou_internado,
29      CAST(B006 AS INT64) AS foi_entubado,
30      CAST(B007 AS INT64) AS possui_planosaude,
31      CAST(B009B AS INT64) AS resultado_swab,
```

```
Consulta_Filtrada EXECUTAR SALVAR CONSULTA FAZER O DOWNLOAD CO

32      CAST(B009D AS INT64) AS resultado_sangue_furodedo,
33      CAST(B009F AS INT64) AS resultado_sangue_veiabraco,
34      CASE
35          WHEN B009B='1' OR B009D='1' OR B009F='1' THEN '1'
36          ELSE '0'
37      END AS positivo_covid,
38      CAST(B0101 AS INT64) AS diabetes,
39      CAST(B0102 AS INT64) AS hipertensao,
40      CAST(B0103 AS INT64) AS doenca_respiratoria,
41      CAST(B0104 AS INT64) AS doencas_coracao,
42      CAST(B0105 AS INT64) AS depressao,
43      CAST(B0106 AS INT64) AS cancer,
44      CASE
45          WHEN B0101='1' OR B0102='1' OR B0103='1' OR B0104='1' OR B0105='1' OR B0106='1' THEN '1'
46          ELSE '0'
47      END AS possui_comorbidade,
48      CAST(B011 AS INT64) AS medidas_restricao,
49      CAST(C001 AS INT64) AS trabalhou,
50      CAST(C007B AS INT64) AS clt_servidor,
51      CAST(C01011 AS INT64) AS faixa_rendimento_habitual,
52      CAST(C01012 AS INT64) AS rendimento_habitual,
53      CAST(C011A11 AS INT64) AS faixa_rendimento_mesvigente,
54      CAST(C011A12 AS INT64) AS rendimento_mesvigente,
55      CAST(E001 AS INT64) AS solicitou_emprestimo,
56      CAST(F002A1 AS INT64) AS possui_sabao_detergente,
57      CAST(F002A2 AS INT64) AS possui_alcool,
58      CAST(F002A3 AS INT64) AS possui_mascara,
59      CAST(F002A4 AS INT64) AS possui_luvas,
60      CAST(F002A5 AS INT64) AS possui_desinfetante_cloro
61 FROM `basedosdados.br_ibge_pnad_covid_microdados`
62 WHERE mes IN(7,8,9)
63 ORDER BY ano, mes
```

ANÁLISE DE DADOS

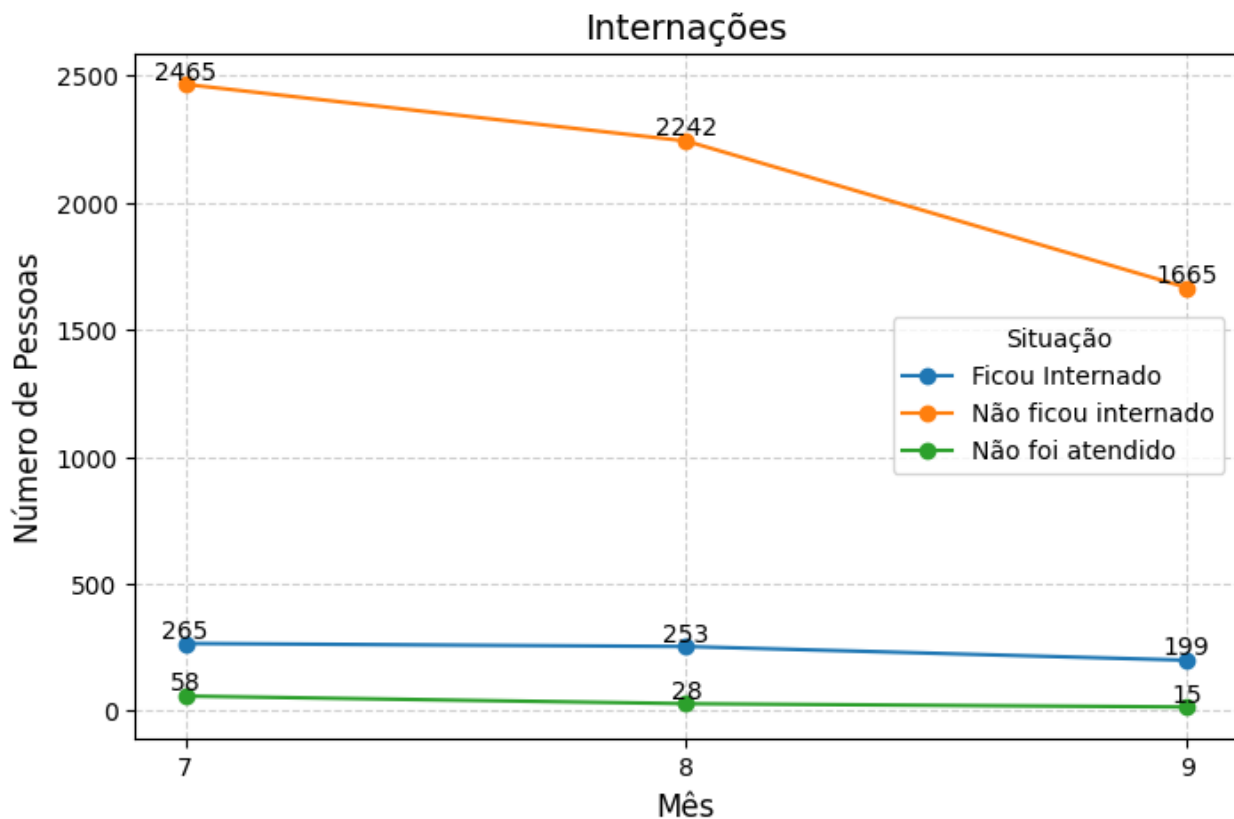
Os dados que serão apresentados a seguir, foram tratados e processados em Python, através do Google Colaboratory.

O código do notebook pode ser acessado em quaisquer um dos links abaixo:

- [Google Colab](#)
- [GitHub](#)

CARACTERIZAÇÃO DOS SINTOMAS CLÍNICOS

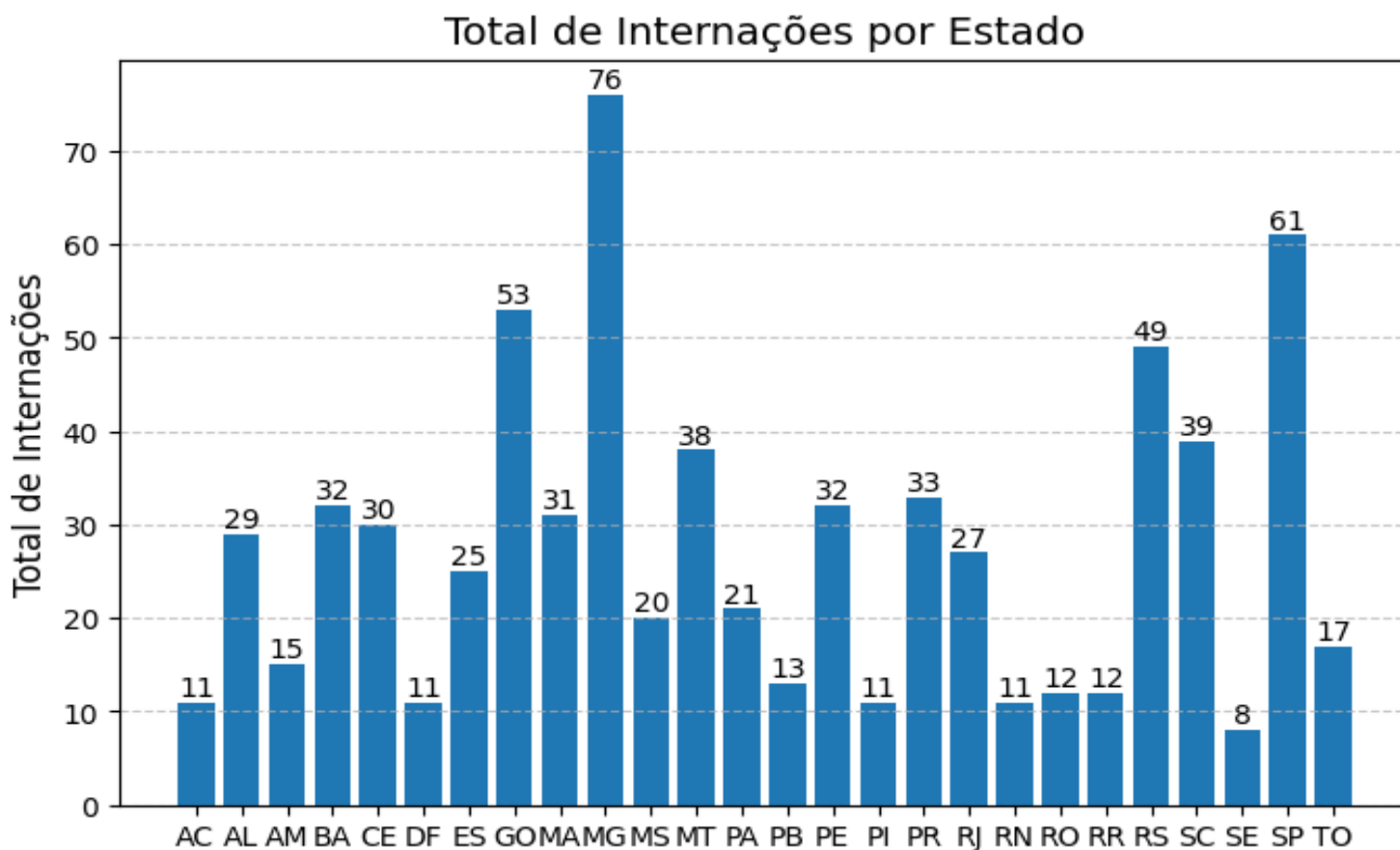
Número de internações entre Julho/20 e Setembro/20



Das pessoas que buscaram auxílio médico, a maior parte dos entrevistados não foram internados. Ao longo dos três meses, podemos perceber uma redução na quantidade de pessoas que buscaram um auxílio profissional.

CARACTERIZAÇÃO DOS SINTOMAS CLÍNICOS

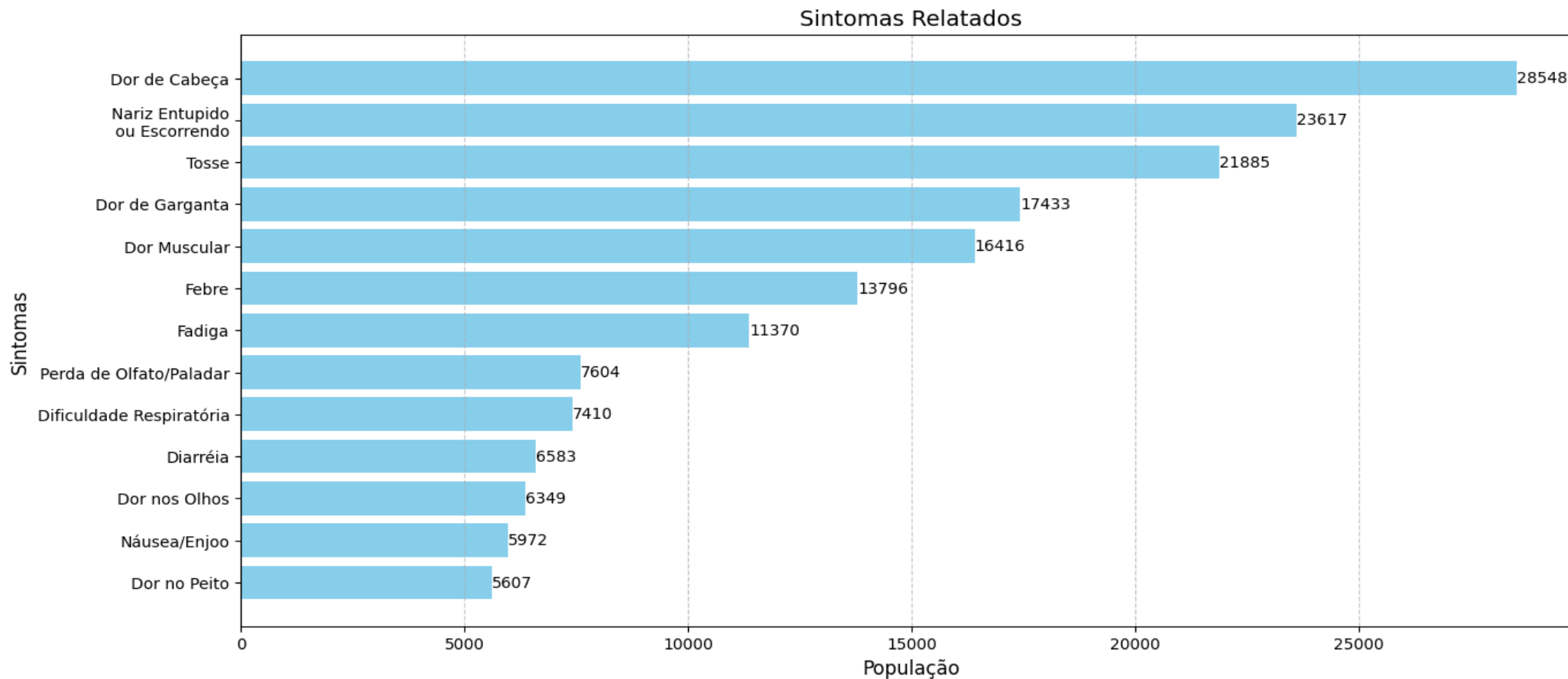
Internações por Unidade Federativa entre Julho/20 e Setembro/20



A região com maior número de internações foi a Nordeste, com 27,9% dos casos. Na sequência, temos a Sudeste com 26,8%.

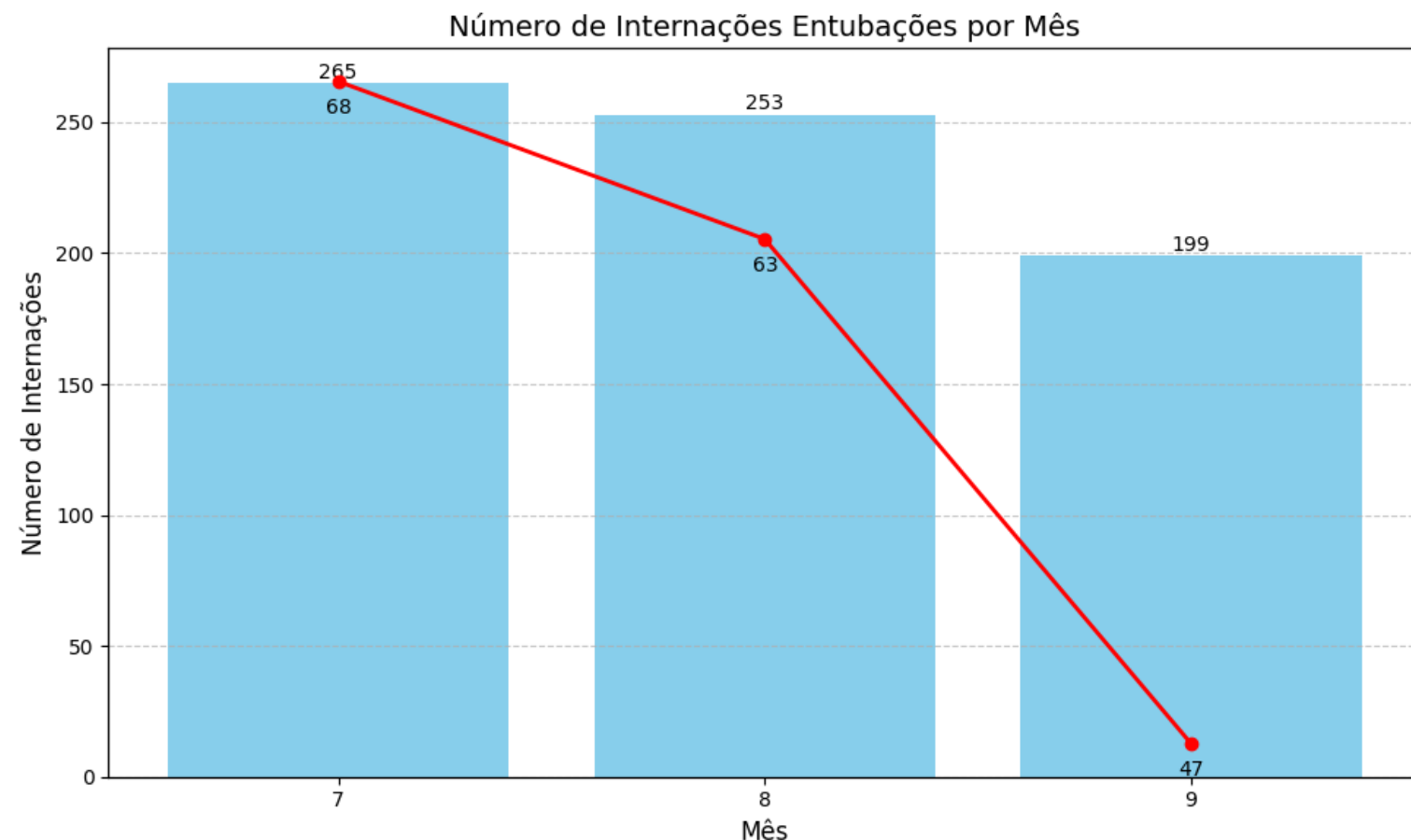
CARACTERIZAÇÃO DOS SINTOMAS CLÍNICOS

Sintomas apresentados



CARACTERIZAÇÃO DOS SINTOMAS CLÍNICOS

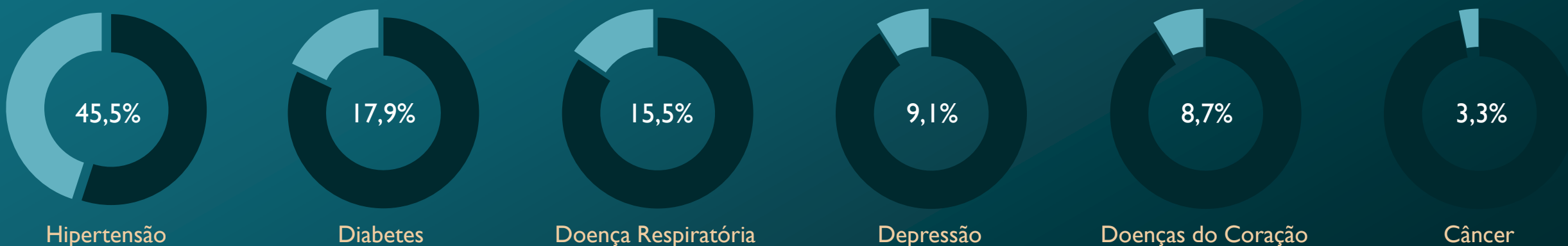
Número de Internações e entubações entre Julho/20 e Setembro/20



A proporção entre internações e entubações apresentou uma leve queda ao longo dos 03 meses. Enquanto 25,7% dos internados foram entubados em Julho/20, a taxa caiu para 23,6% em Setembro/20.

CARACTERIZAÇÃO DOS SINTOMAS CLÍNICOS

Comorbidades



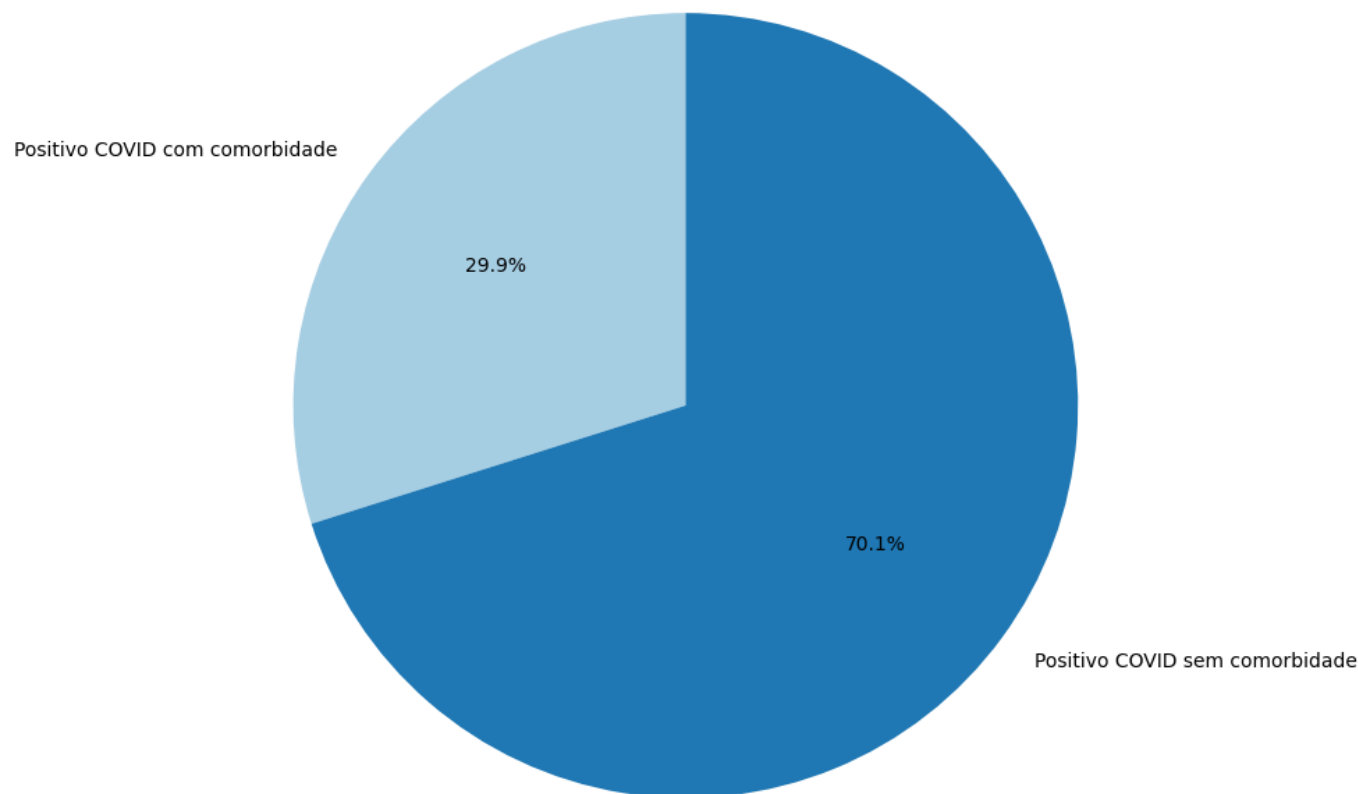
Dos participantes da pesquisa, 45,5% apresentava hipertensão.

Outro dado alarmante é a proporção de participantes com depressão, que apesar de não ser a mais prevalente, poderia ter seus sintomas acentuados com as medidas de restrições necessárias durante a pandemia.

CARACTERIZAÇÃO DOS SINTOMAS CLÍNICOS

Número de Internações e entubações entre Julho/20 e Setembro/20

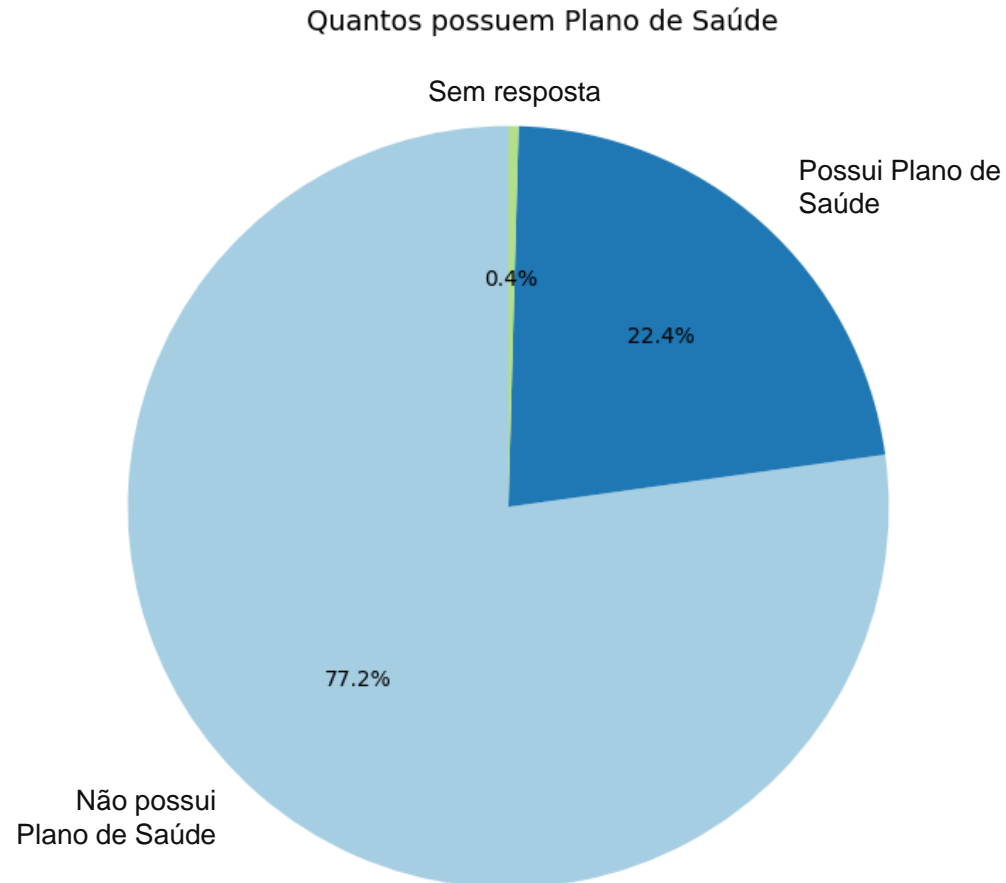
Distribuição de Casos Positivos de COVID



De todos os casos entrevistados, 70,1% que testaram positivo para a COVID-19 não possuíam comorbidades.

CARACTERIZAÇÃO DOS SINTOMAS CLÍNICOS

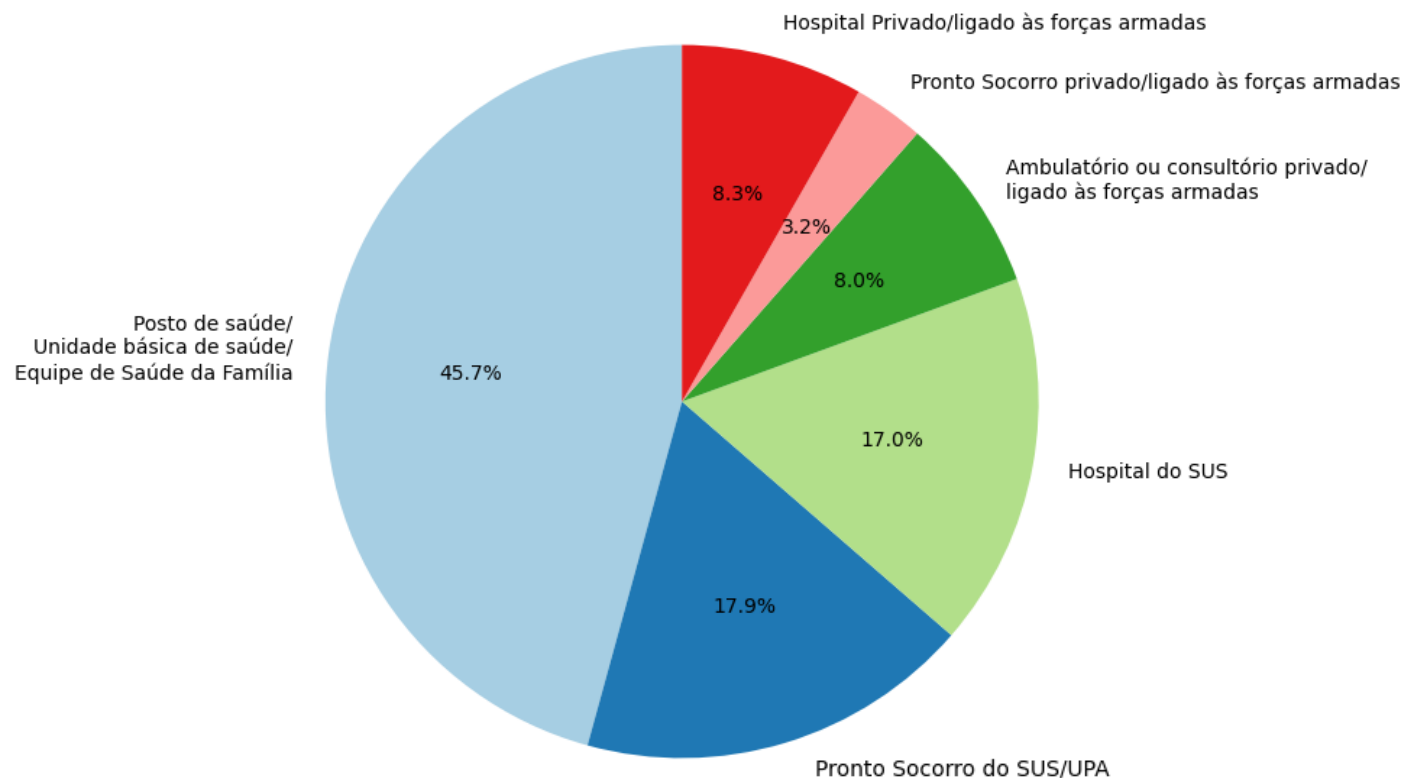
População com Plano de Saúde



CARACTERIZAÇÃO DOS SINTOMAS CLÍNICOS

Estabelecimento de Internação

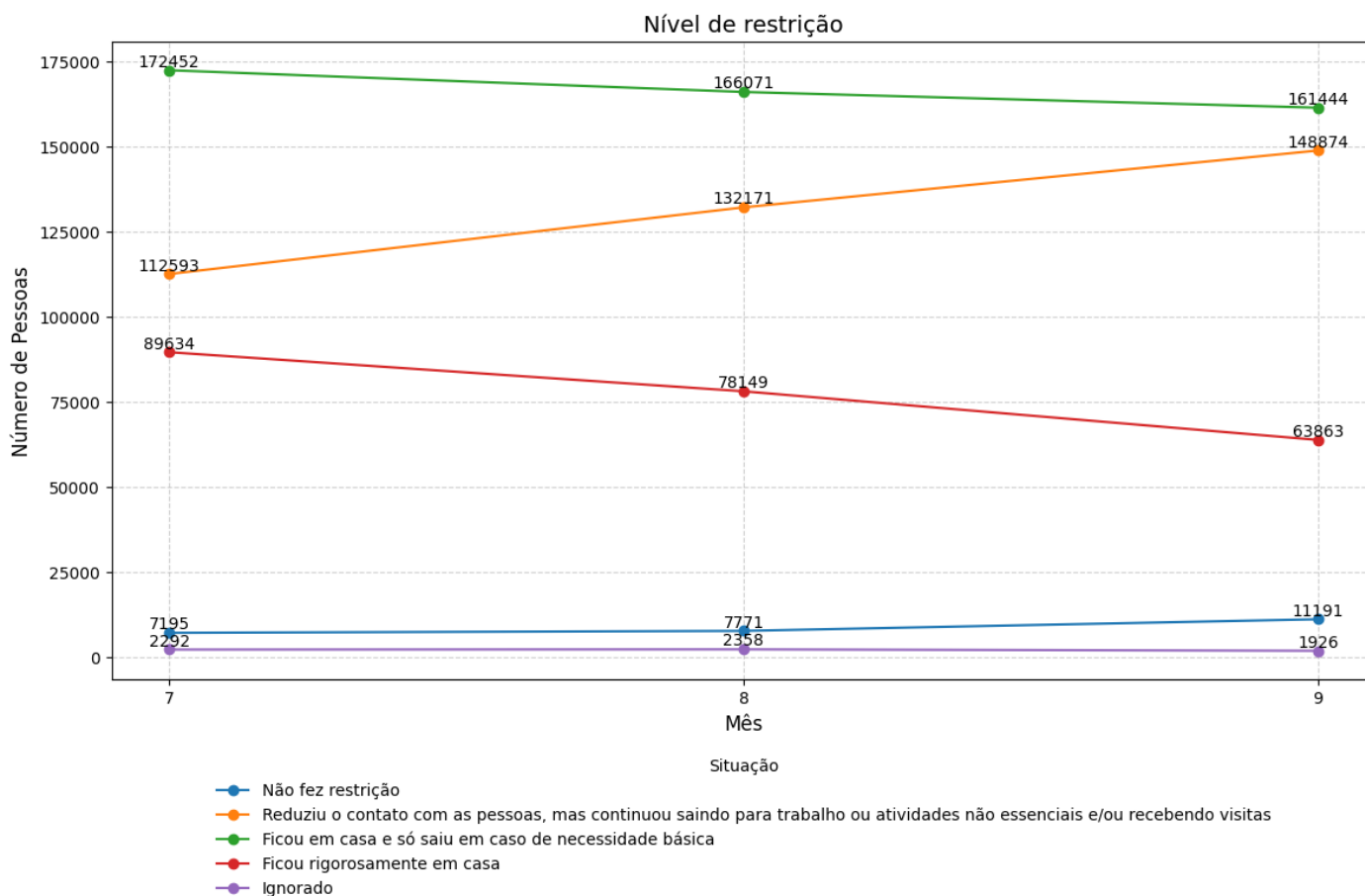
Local em que buscou atendimento



Mais de 80% da população entrevistada recorre ao SUS para ter um atendimento, seja em Postos de Saúde, UPA/PS ou Hospital.

COMPORTAMENTO DA POPULAÇÃO NA ÉPOCA DA COVID

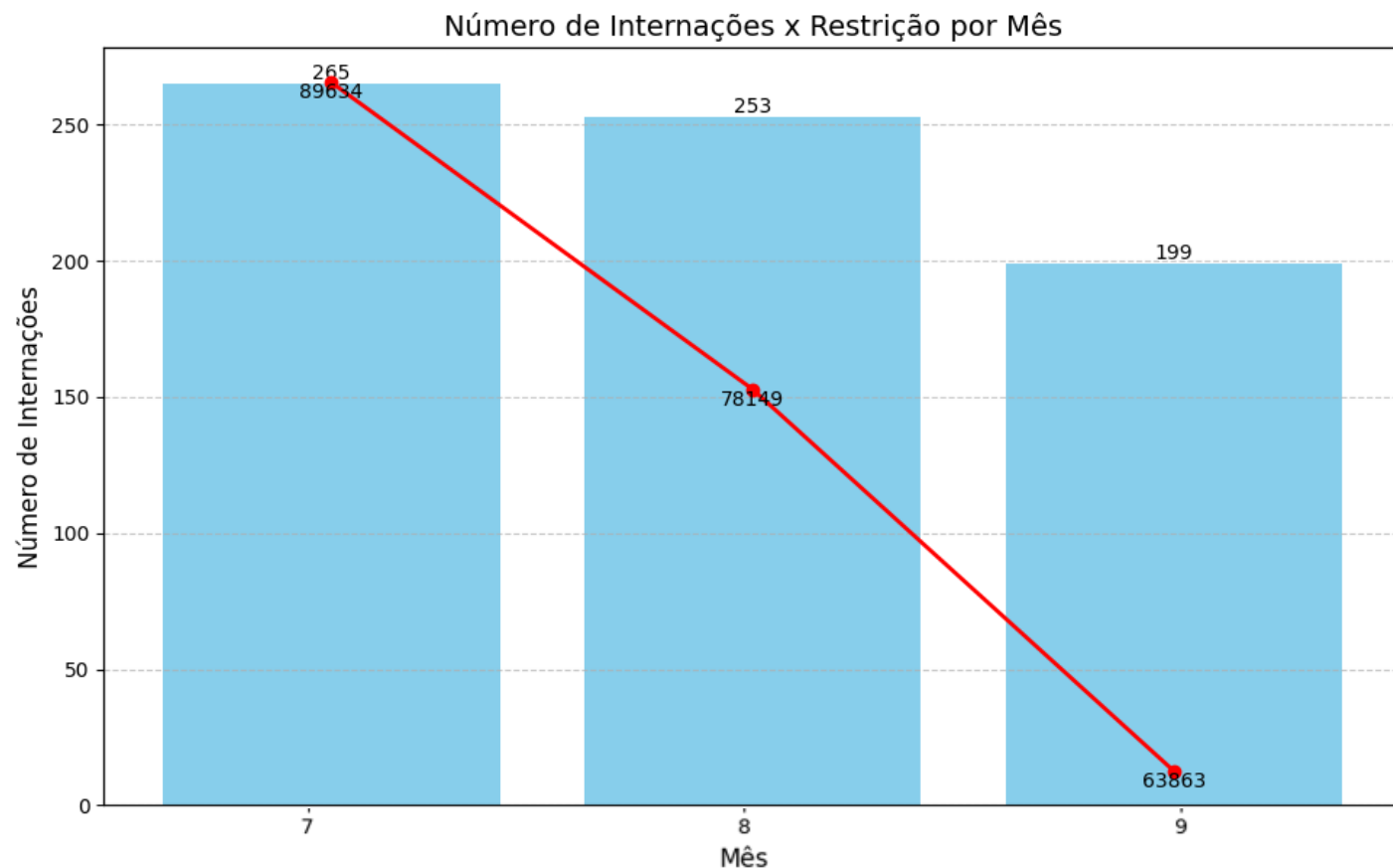
Medidas de Restrição



Pelo gráfico, vemos que ao passar dos meses as pessoas que estavam fazendo restrição mais rigorosa ou saindo apenas para necessidade básica, passaram a sair de casa para trabalhar ou desenvolver atividades não-essenciais.

COMPORTAMENTO DA POPULAÇÃO NA ÉPOCA DA COVID

Relação entre número de internações e medidas rigorosas de restrição

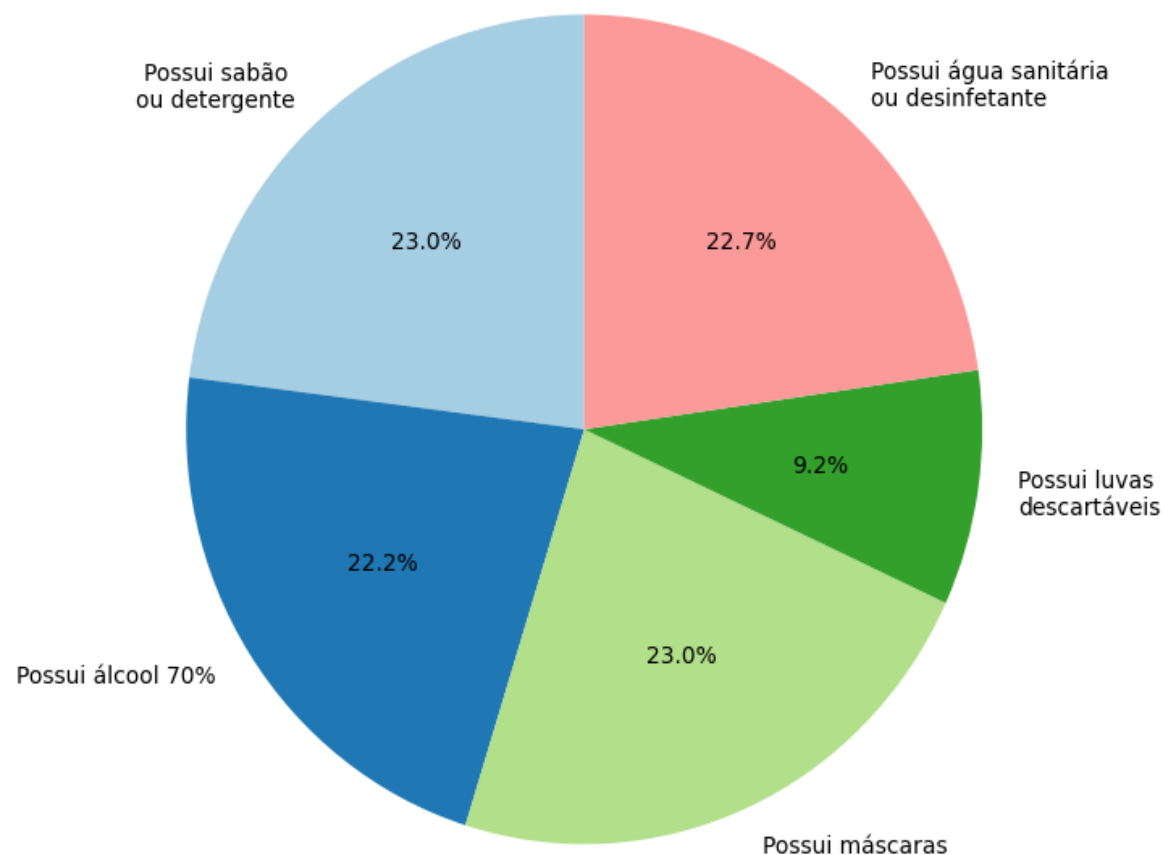


Percebemos uma queda no número pessoas que adotaram medidas de restrições mais rigorosas, sem sair de casa, na medida em que as internações também diminuíram ao longo dos meses.

COMPORTAMENTO DA POPULAÇÃO NA ÉPOCA DA COVID

Itens de higiene

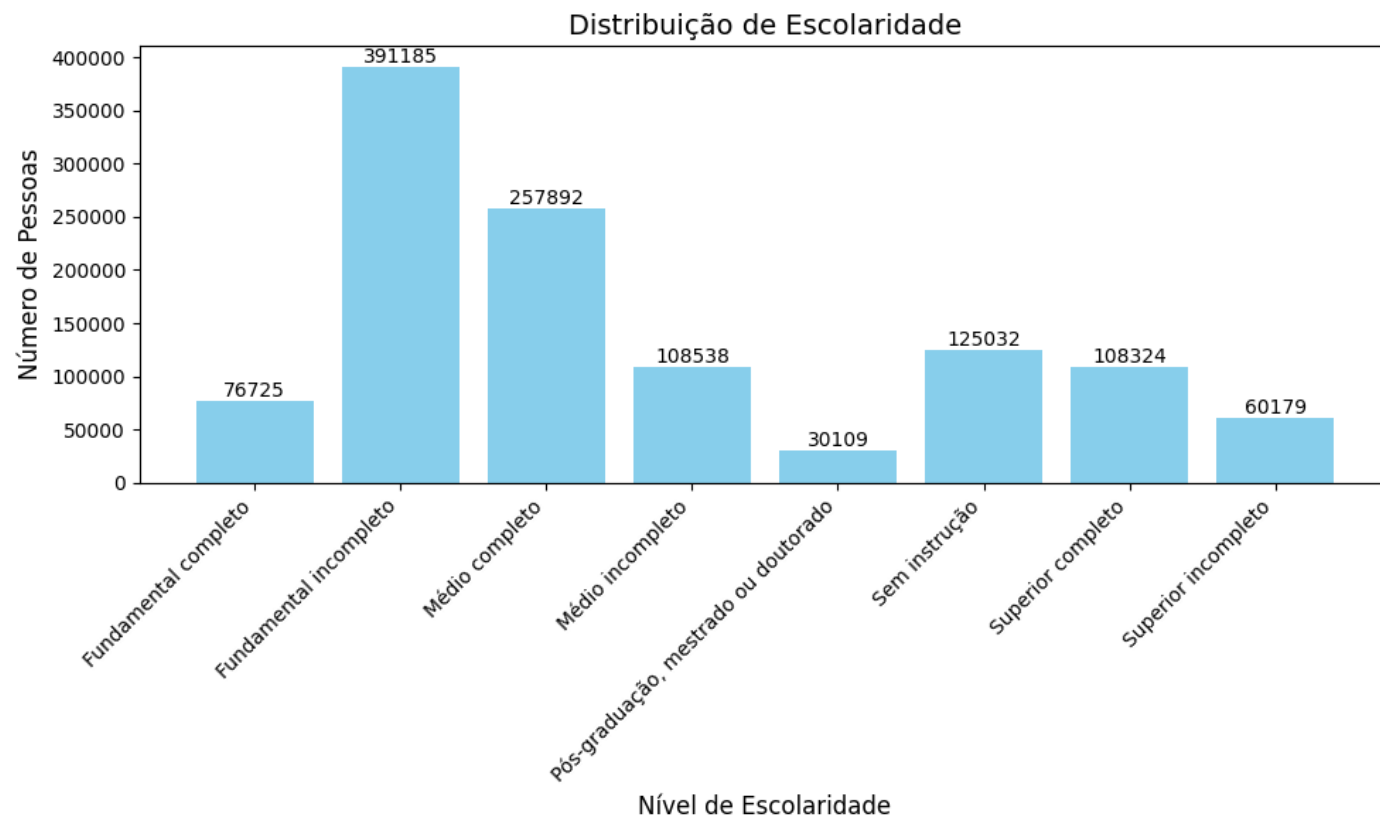
Representatividade dos produtos de higiene



Com exceção das luvas descartáveis, os demais itens de higiene possuíam prevalência semelhante na residência dos entrevistados.

CARACTERÍSTICAS ECONÔMICAS DA SOCIEDADE

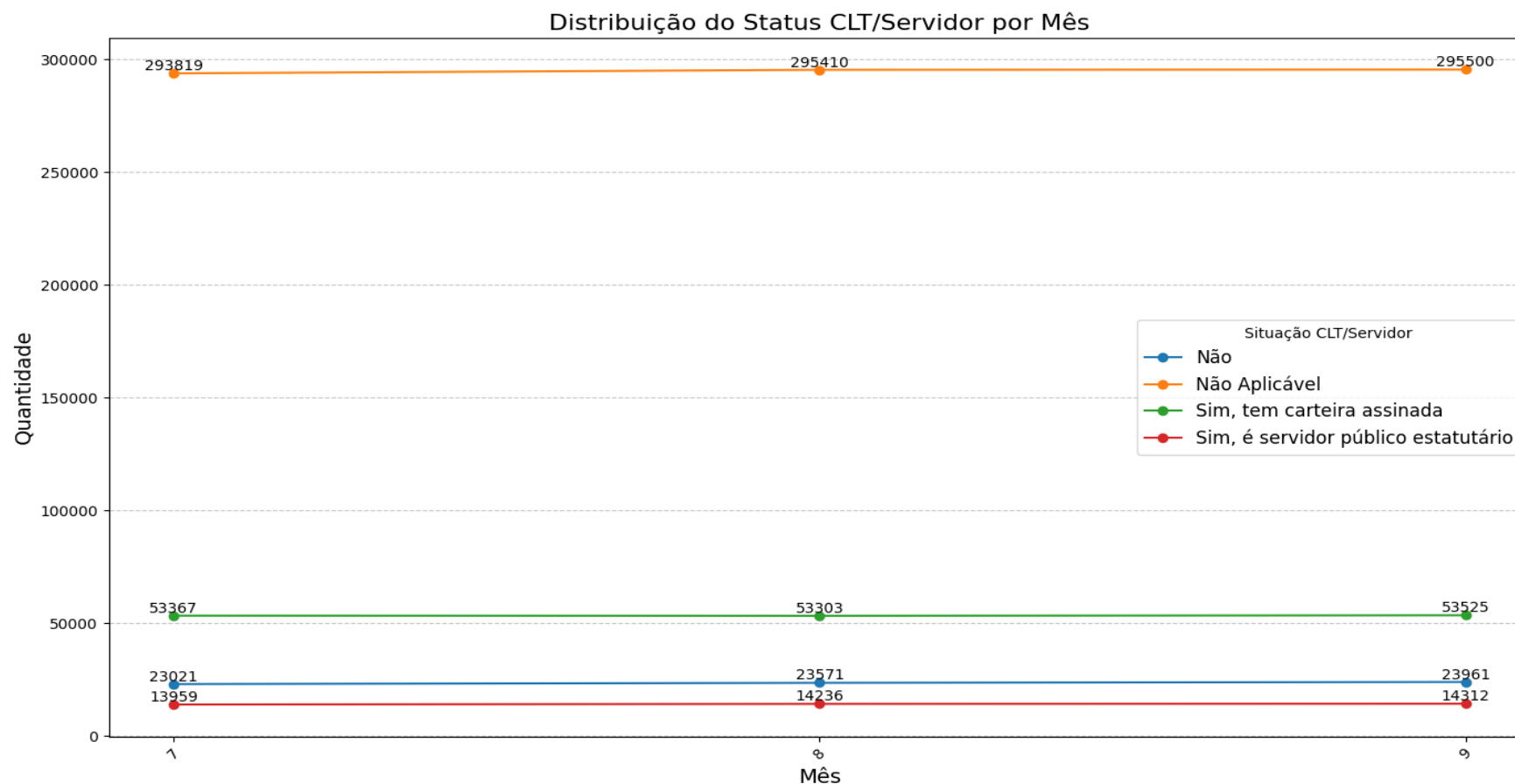
Escolaridade



Analisando todas as respostas obtidas nos três meses, 34% possui o fundamental incompleto. Em contrapartida, apenas 3% possui pós-graduação.

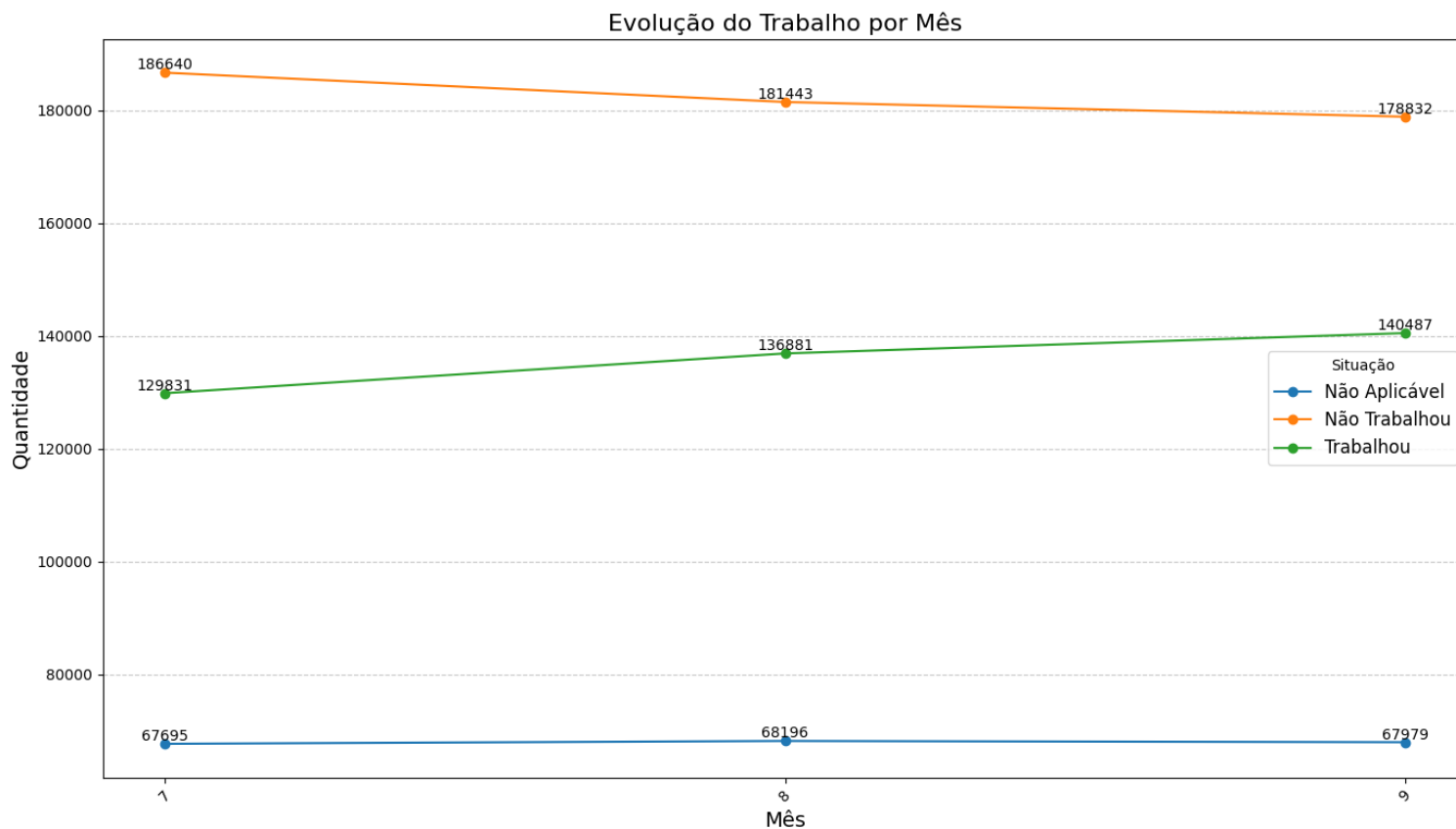
CARACTERÍSTICAS ECONÔMICAS DA SOCIEDADE

População com carteira assinada CLT e servidores públicos no período



CARACTERÍSTICAS ECONÔMICAS DA SOCIEDADE

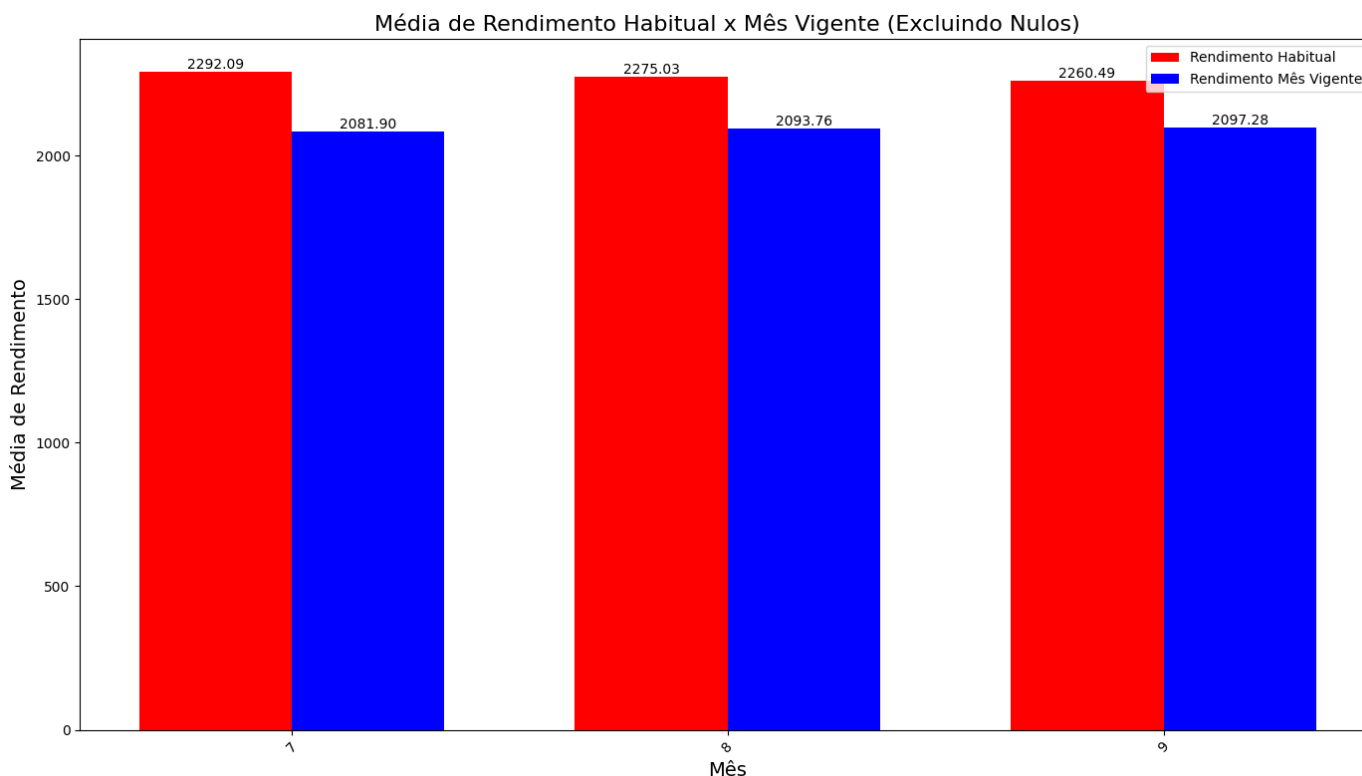
Trabalho no período entre Julho/20 e Setembro/20



Com a diminuição das interações, tivemos o movimento de retomada das atividades, tendo um leve aumento ao longo dos meses no número de pessoas que trabalharam, e em contrapartida, diminuição daqueles que não trabalharam.

CARACTERÍSTICAS ECONÔMICAS DA SOCIEDADE

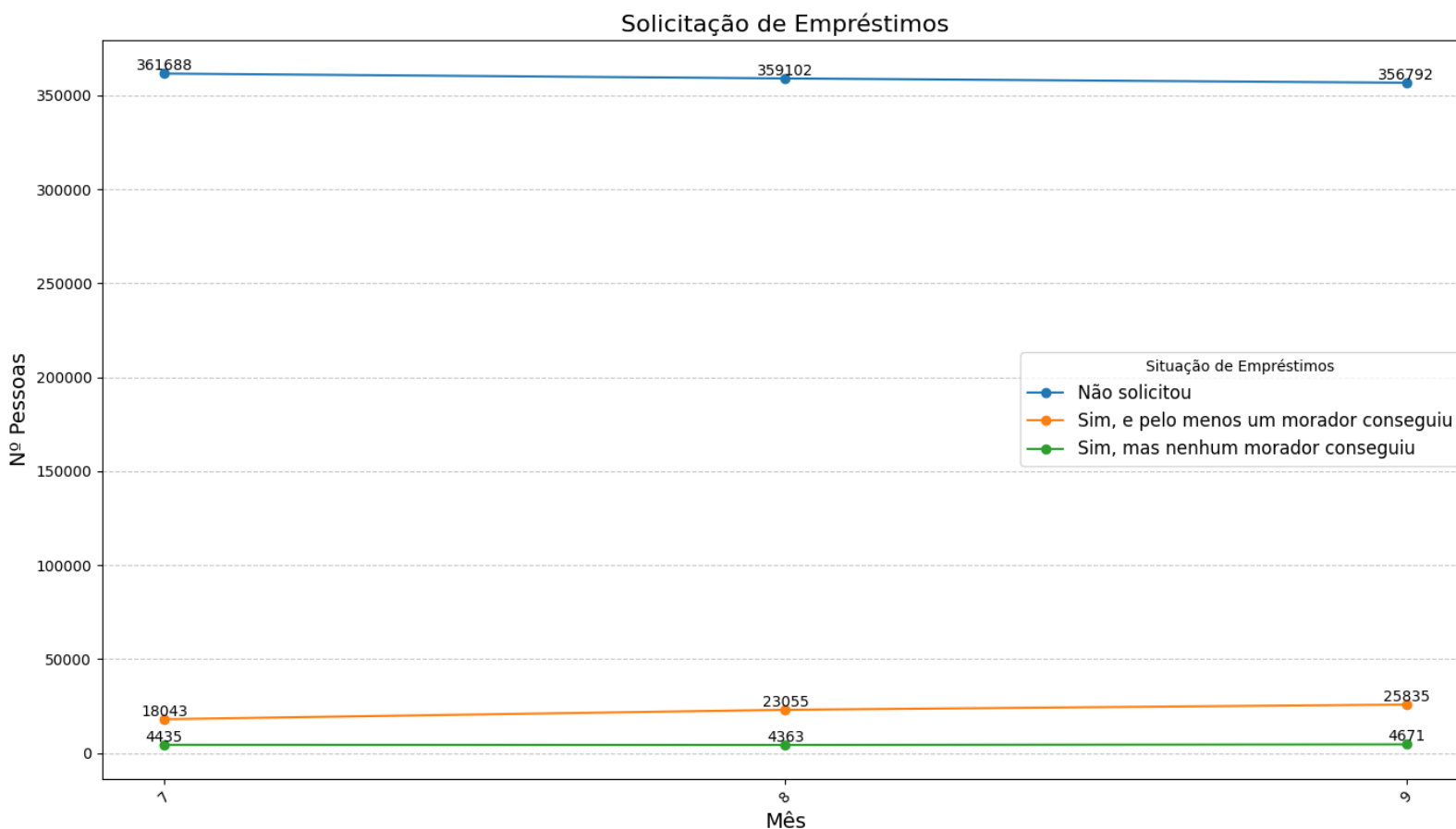
Comparação entre rendimento habitual e vigente



Comparando a média do rendimento bruto habitual e o recebido ao longo dos três meses, considerando apenas aquelas pessoas que receberão algum valor, percebemos que a pandemia impactou nos valores esperados. A redução média foi de - 8,12%.

CARACTERÍSTICAS ECONÔMICAS DA SOCIEDADE

Solicitação de empréstimo no período



Apesar do movimento de retomada das atividades, ocorreu um aumento da população que solicitou empréstimo ao longo dos meses.

CONCLUSÃO

Mesmo utilizando apenas 03 meses para a análise dos dados, é possível obtermos informações importantes para entender características da população e assim reduzir danos em caso de surtos de novas doenças.

Pudemos notar que a região Nordeste foi a que mais teve internações no período analisado, seguido da Sudeste, indicando a importância de ampliar os leitos disponíveis nas regiões. Vale ressaltar que o cenário nacional indica que em média, 25% foram entubados, aumentando o custo da internação e os riscos atrelados.

A população sofreu impactos financeiros no período, reduzindo a renda bruta e aumentando a busca por empréstimos pessoais. Além disso, uma parcela pequena da população possui plano de saúde, reforçando a necessidade de investimento no SUS e a importância que o mesmo teve durante o período da pandemia.

Além dessas medidas, é fundamental a disseminação de informação para a população, estimulando medidas de restrição e higiene.



FIAP

Flora Serafim de Carvalho



flora_scarvalho@live.com



RM: 354000