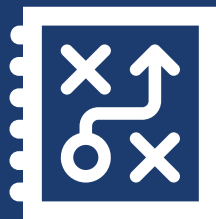




TECH CHALLENGE

FASE 3 – PÓS TECH DATA ANALYTICS - FIAP

DESAFIO



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.

REQUISITOS



Uso da base de dados PNAD-COVID-19 para análise de comportamento, sintomas clínicos e econômicas da população



Utilização de no máximo 20 questionamentos realizados na pesquisa



Utilização de uma base de 03 meses para construção da solução

OBJETIVOS



Efetuar organização dos dados e realizar uma breve análise



Ações necessárias em caso de um novo surto de COVID-19

PASSOS



Etapa 1

Análise da base disponibilizada no site do IBGE para compreensão dos dados com uso dos dicionários disponibilizados.



Etapa 2

Carregamento dos dados do IBGE na plataforma Google Cloud (GCP).



Etapa 3

Criação das tabelas dimensões para estruturação e análise dos dados.



Etapa 4

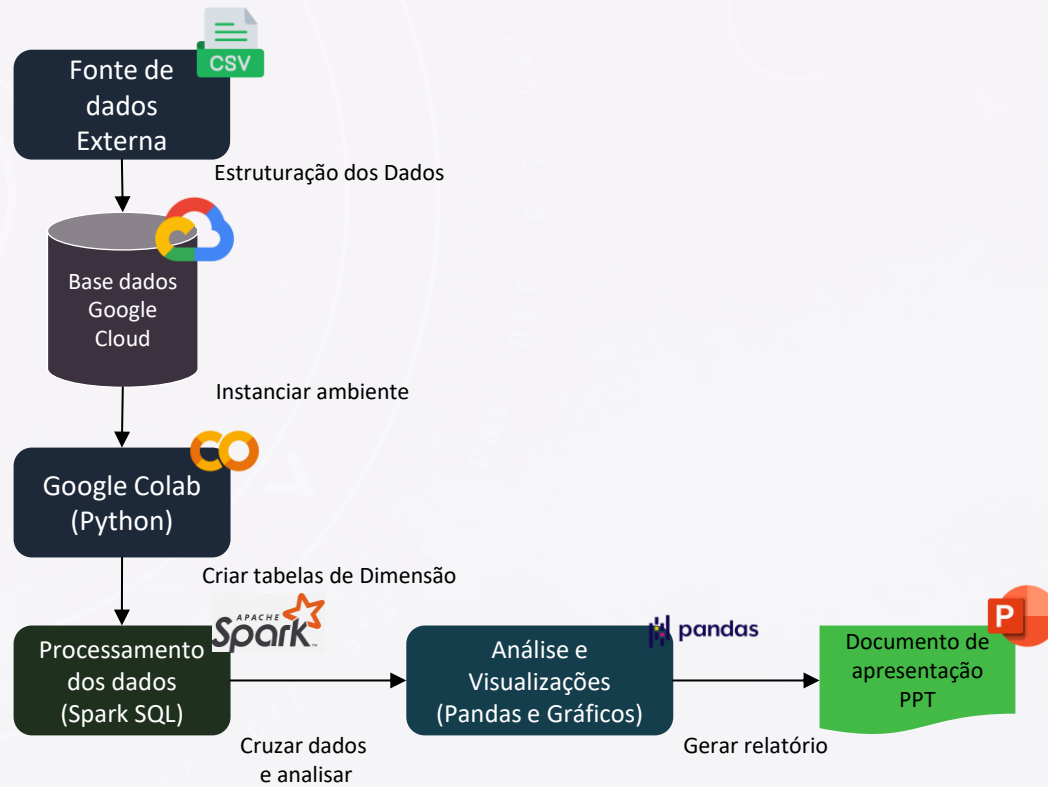
Aplicadas técnicas de transformação de dados e cruzamentos com as tabelas dimensões.



Etapa 5

Geração de insights através de análises e geração de gráficos para visualização.

Arquitetura do processo

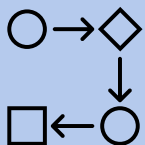


Ferramentas Google

Foram utilizadas diversas ferramentas da Google para o projeto, sendo elas GCP(Cloud), BigQuery, Cloud Storage e Colab

Ferramentas fase

Foram utilizadas tecnologias apresentadas durante a fase 3, sendo elas google colab, linguagem Pyspark e arquivos formato CSV



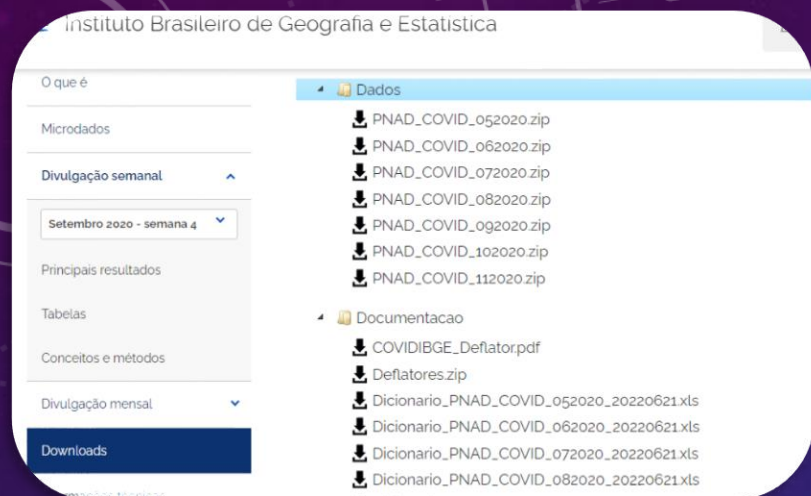
Fluxograma dos dados e ferramentas utilizadas



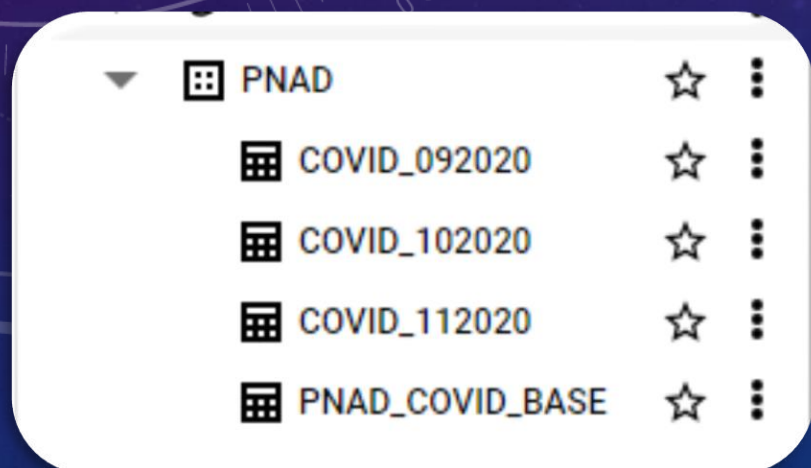
ORGANIZAÇÃO DOS DADOS

ORGANIZAÇÃO DOS DADOS

Arquivos no Site do IBGE



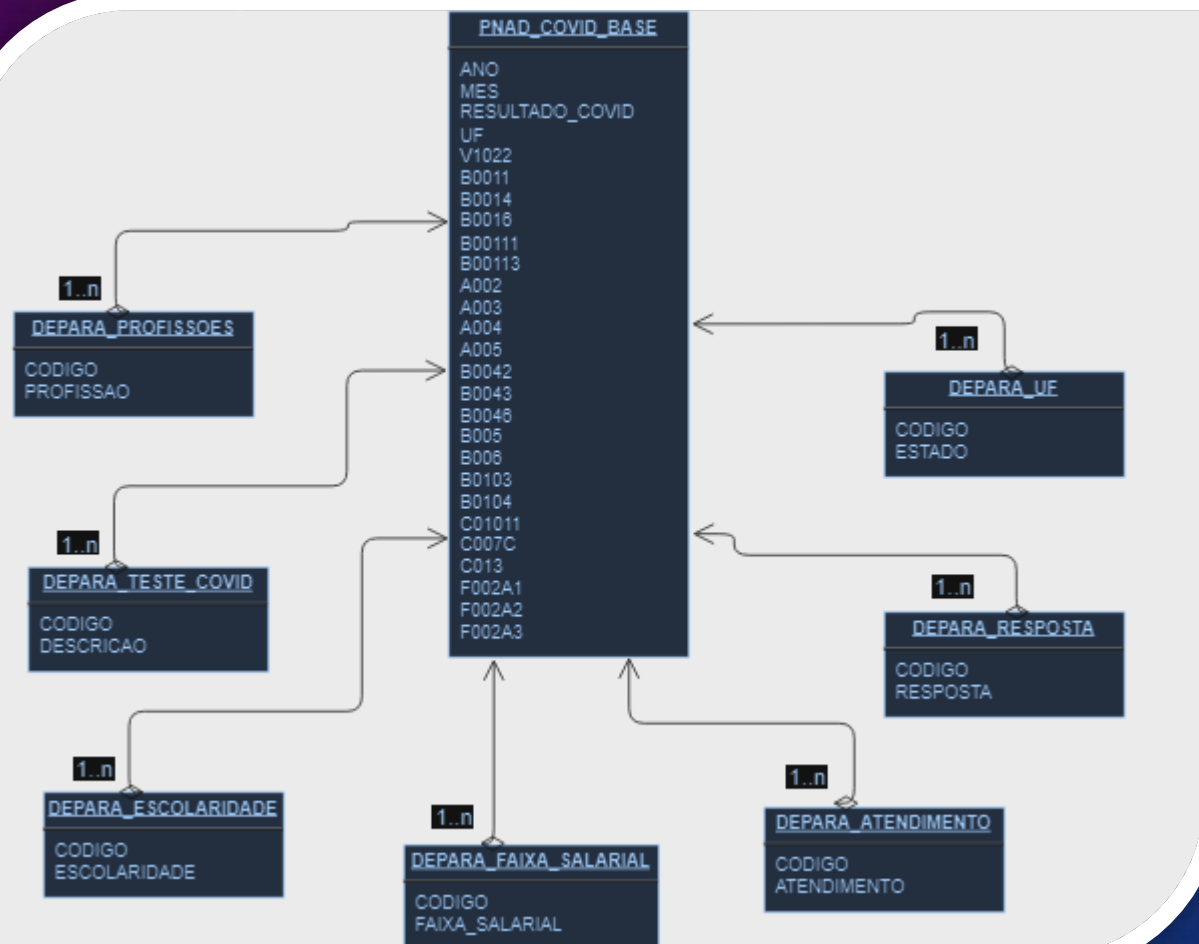
Tabelas Criadas (GCP)



- Organizamos o banco de dados na nuvem (GCP), utilizamos três arquivos .CSV extraídos do IBGE. Cada arquivo contém dados de um mês da pesquisa PNAD, com informações sobre os entrevistados.
- Das 20 questões selecionadas para a análise. Consolidamos os três arquivos em uma única tabela no GCP por meio de uma consulta.
- Ao analisar o dicionário de dados, identificamos três colunas relacionadas a diferentes testes de COVID-19. Decidimos unificá-las em uma coluna, utilizando o melhor resultado de cada caso.
- Por exemplo, se um teste resultou positivo e os outros não foram realizados, registramos apenas a informação "positivo" na nova coluna.

ORGANIZAÇÃO DOS DADOS

Modelo de Entidade e Relacionamento

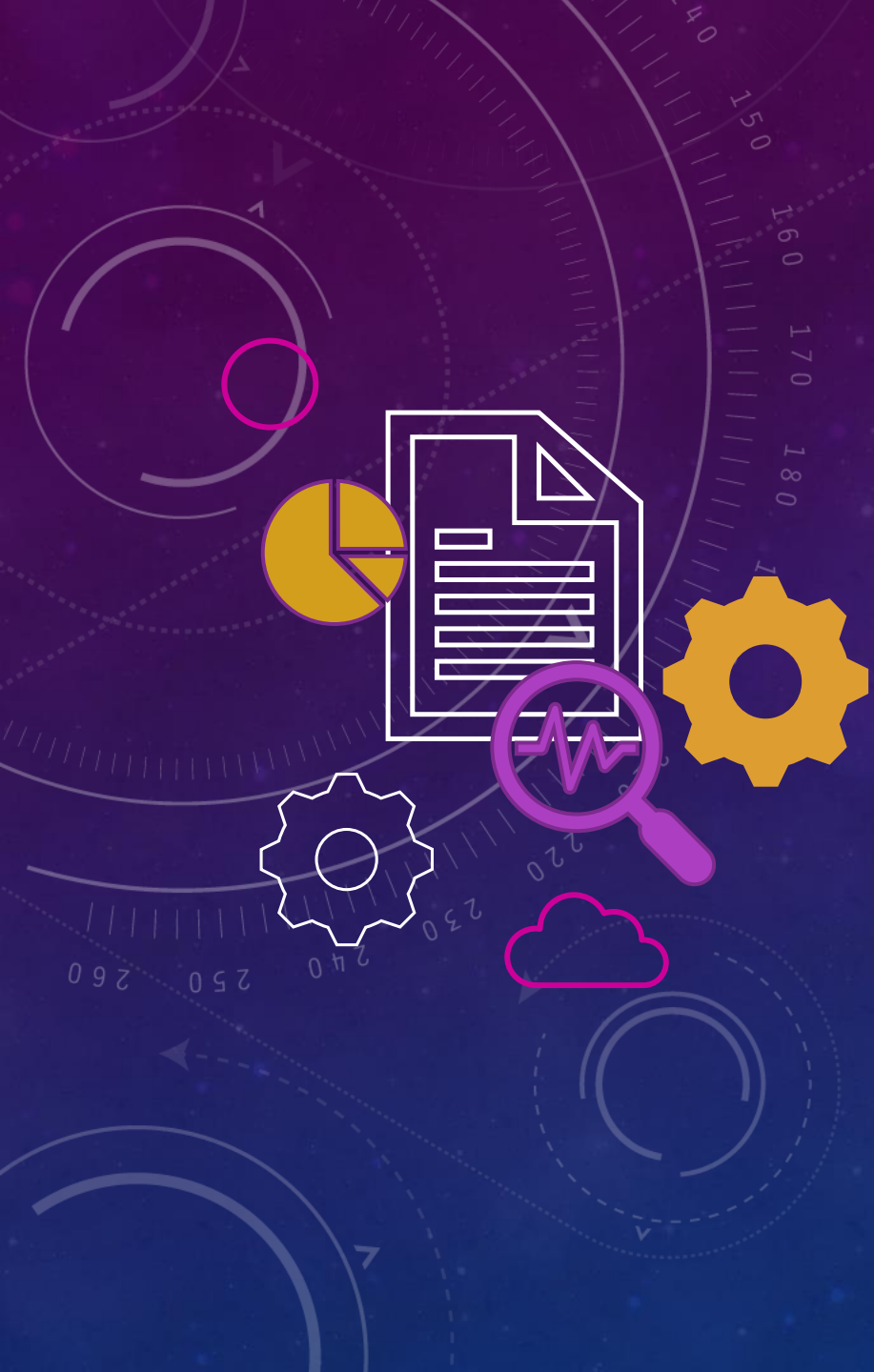


Após configurar o banco de dados na nuvem(GCP), integramos o ambiente ao Google Colab para realizar consultas SQL, aproveitando o poder de processamento do BigQuery.

No entanto, durante as primeiras consultas, percebemos que as colunas do banco não estavam devidamente nomeadas, apresentando apenas os IDs das respostas.

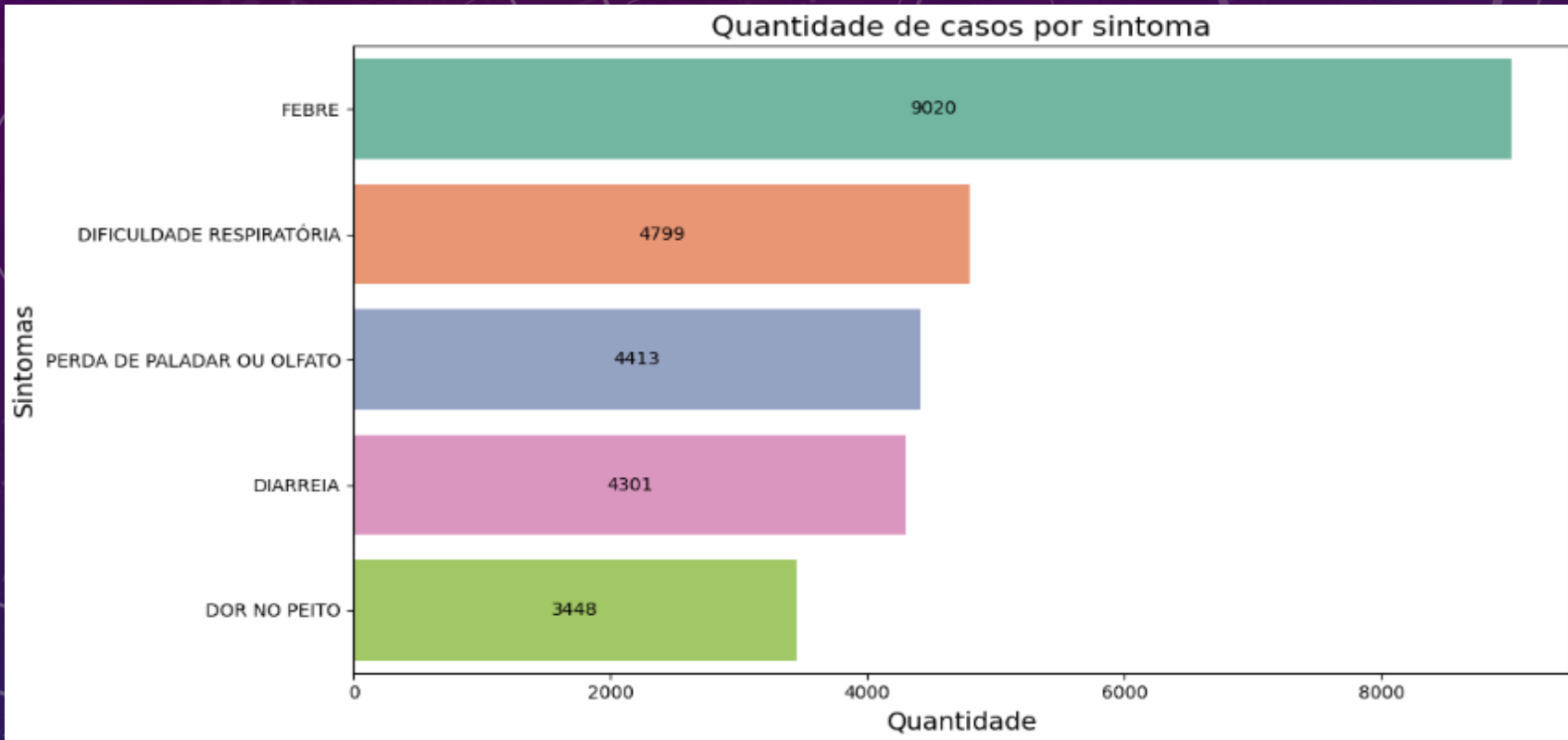
Para resolver esse problema, optamos por criar várias views temporárias utilizando a API do Spark SQL. Essas tabelas foram estruturadas como tabelas de dimensões, permitindo associar os IDs às descrições correspondentes. Em seguida, os resultados das consultas realizadas no BigQuery foram atribuídos às tabelas temporárias no Spark, facilitando o cruzamento dos dados com as novas dimensões.

Esse processo nos permitiu consolidar os dados em Dataframes completos e organizados, prontos para a geração de gráficos e obtenção de insights valiosos.



EXPLORAÇÃO DOS DADOS

SINTOMAS



9.020 casos

FEBRE

Durante as análises foi possível observar que a Febre foi o principal sintoma

4.799 casos

DIFICULDADE RESPIRATÓRIA

A dificuldade para respirar foi o segundo sintoma mais identificado com quantidade de casos

4.413 casos

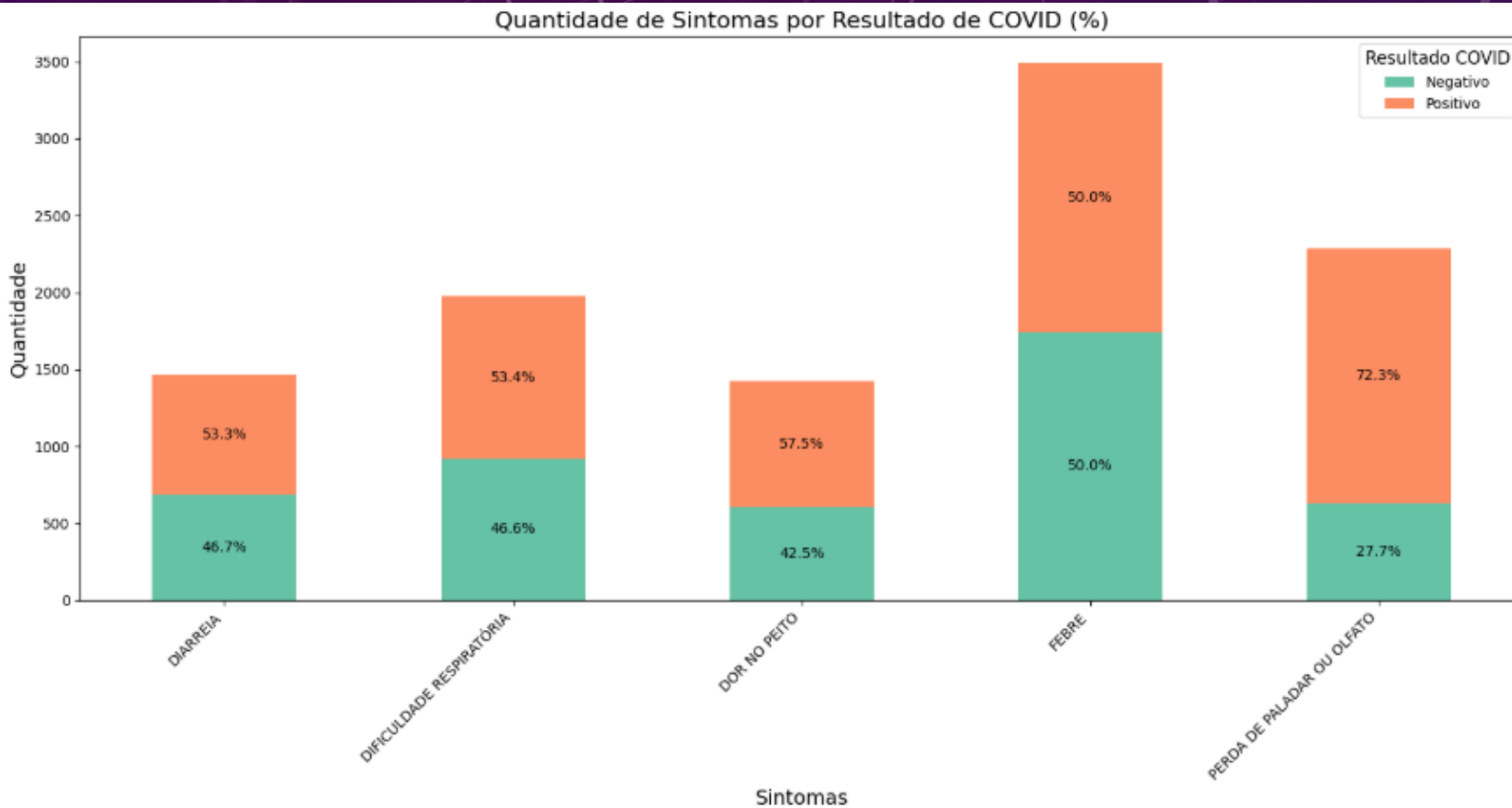
PERDA DO PALADAR

A perda do paladar ou olfato foi a terceira causa, com mais quantidade de casos



Sintomas mais presentes na pesquisa

SINTOMAS



50%

POSITIVO

FEBRE

50% dos entrevistados que foram testados como positivo no teste de Covid tiveram o sintoma

72%

POSITIVO

PERDA DE PALADAR OU OLFATO

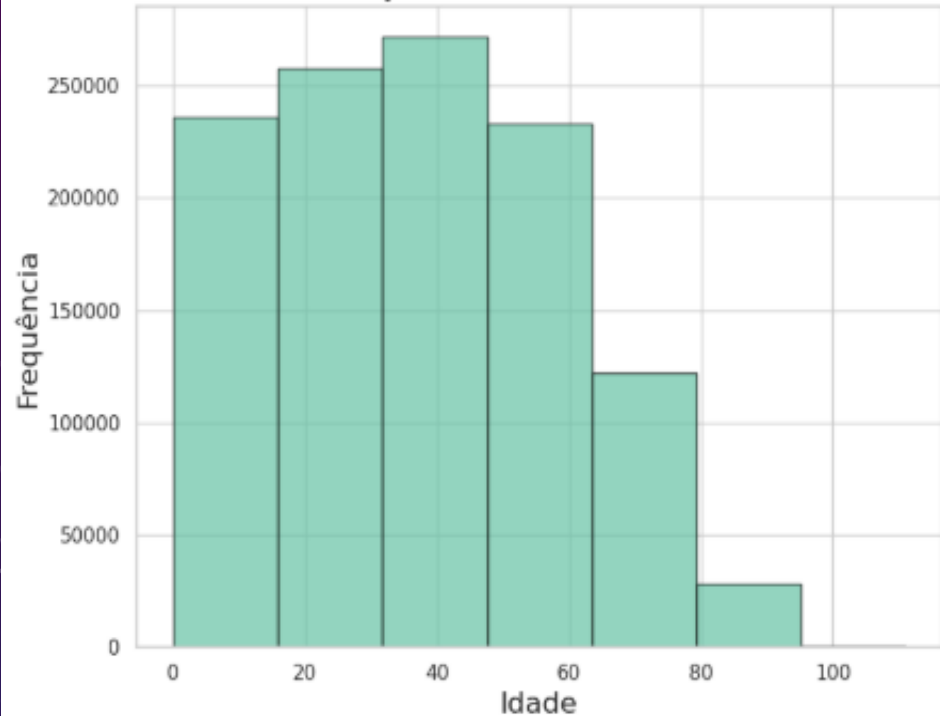
72% dos entrevistados que foram testados como positivo no teste de Covid tiveram o sintoma



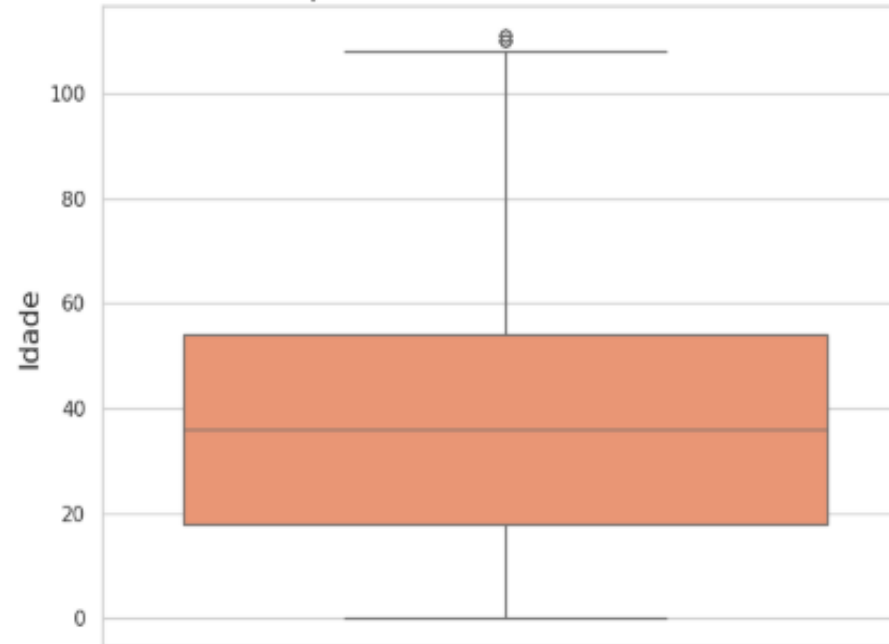
Casos com resultados de testagem para COVID-19, positivos e negativos, associados à presença de algum sintoma clínico.

Entrevistados

Distribuição da idade dos Entrevistados



Boxplot da idade dos Entrevistados



20 e 60 +



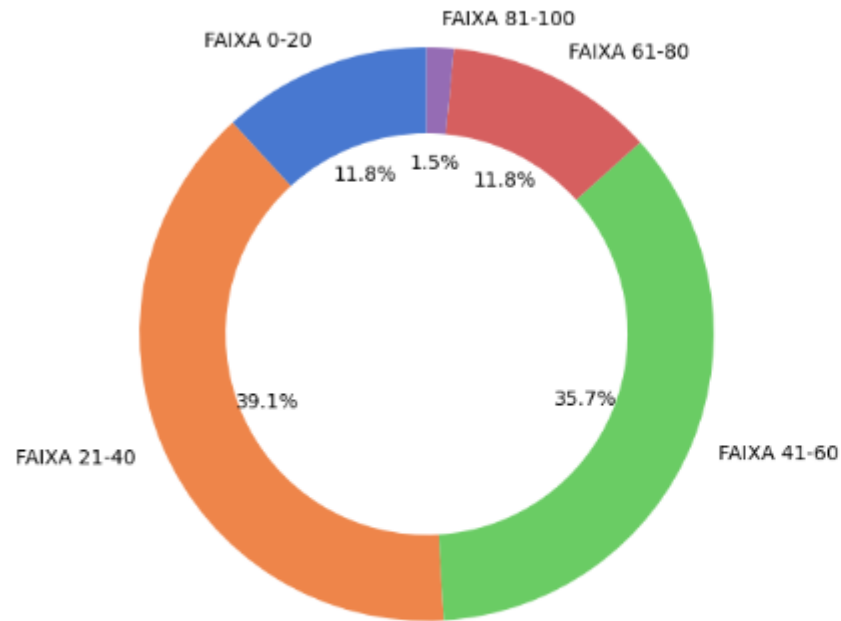
A grande parte dos entrevistados têm a idade entre 20 e 60 anos



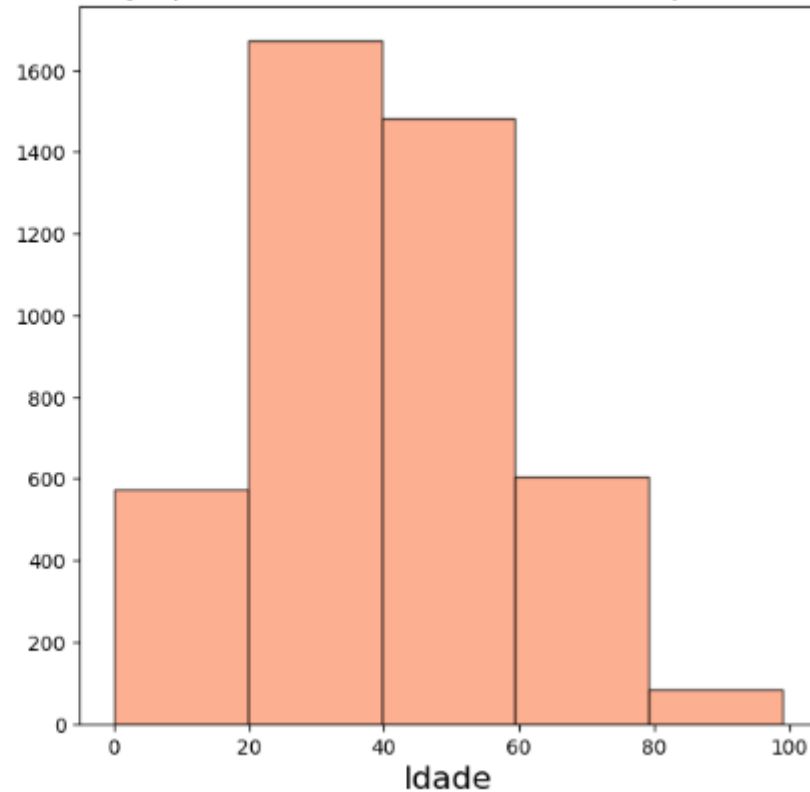
Faixa etária de idade dos entrevistados da pesquisa

Entrevistados

Testes Positivos de Covid por Faixa Etária



Distribuição por idade dos entrevistados com falta de paladar ou olfato



74,8%

Gráfico 1

A faixa etária entre 21 e 60 anos apresentou maior predominância de resultados positivos para COVID-19.

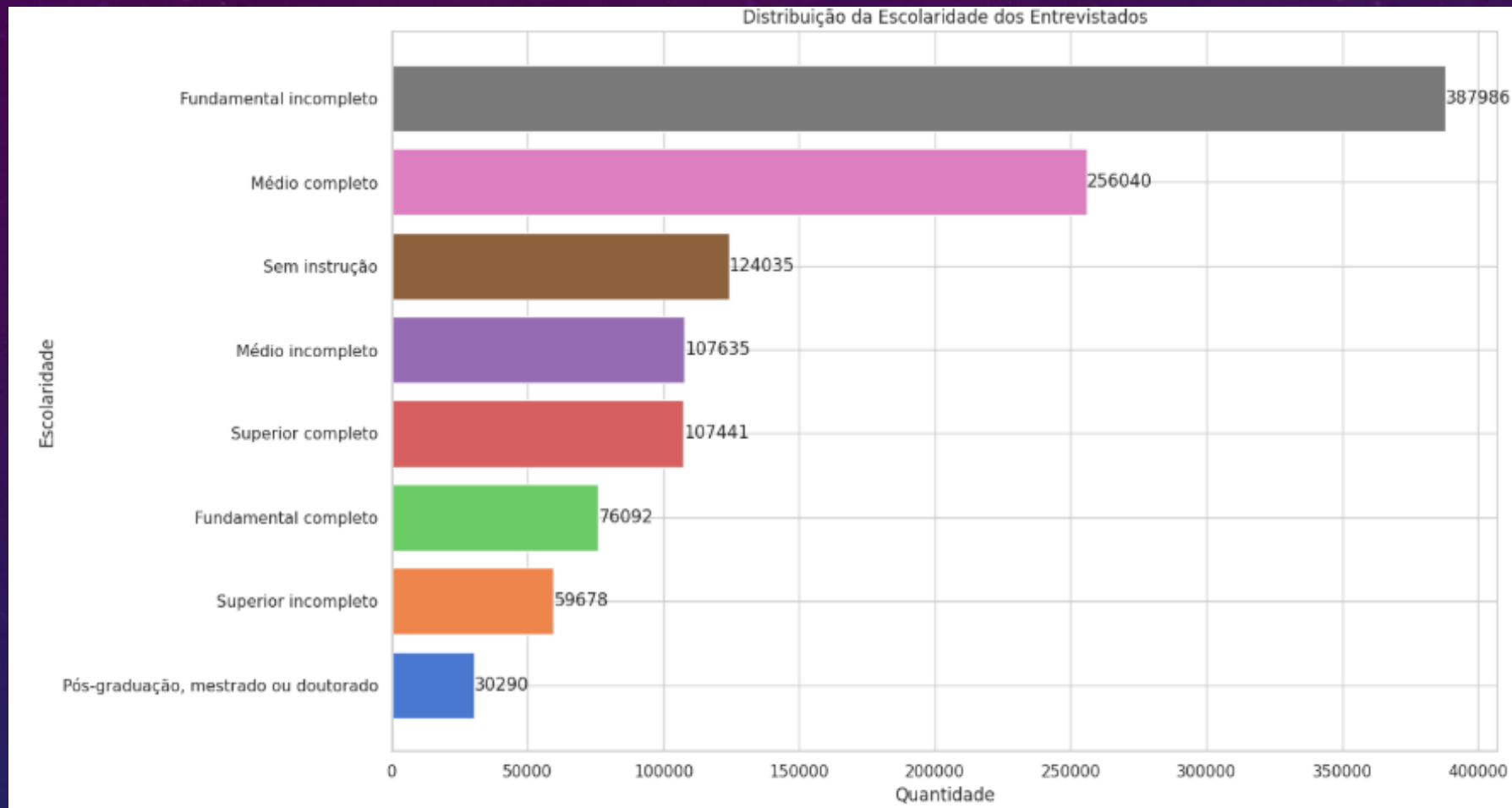
Gráfico 2

Entre os entrevistados que relataram perda de olfato ou paladar, a maioria pertence à faixa etária de 21 a 60 anos.



Análise por faixa etária dos entrevistados

Entrevistados



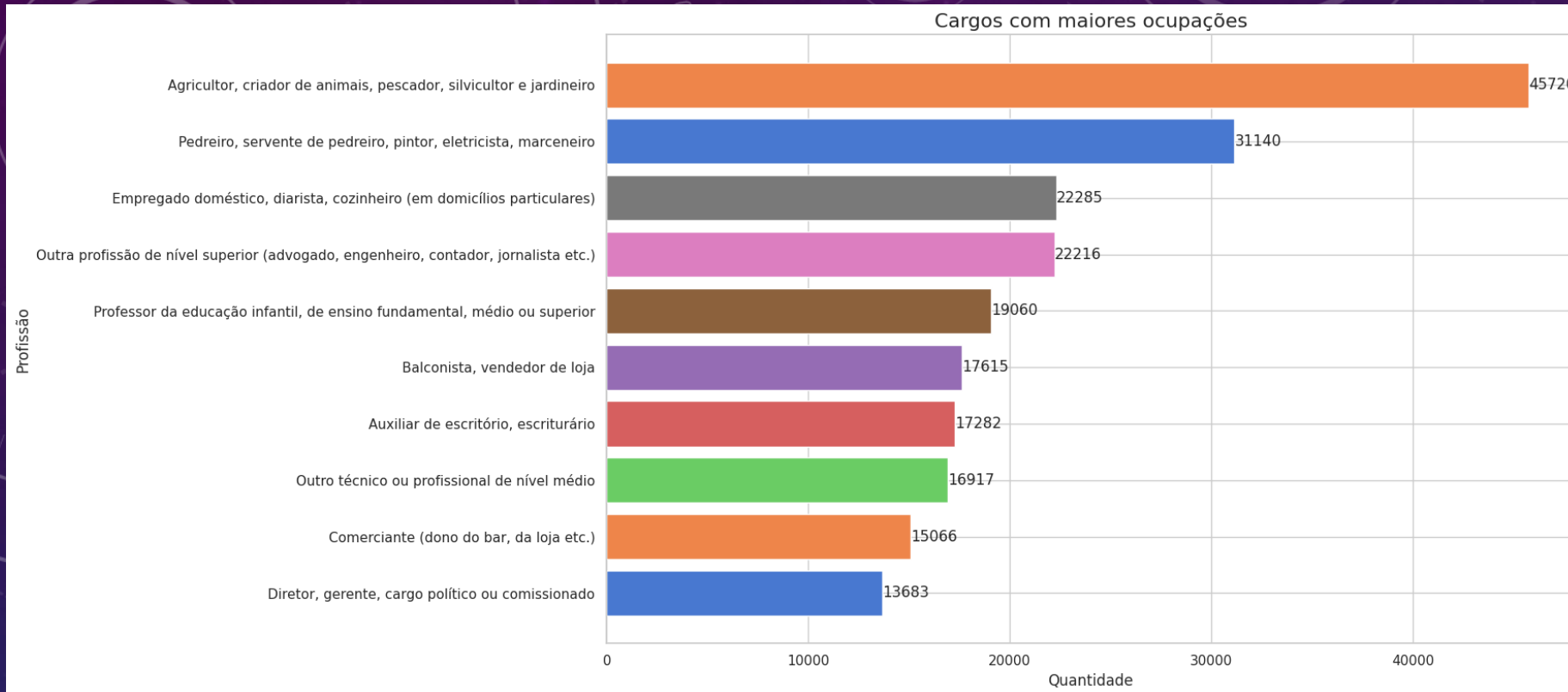
Característica de escolaridade dos entrevistados



População

387.986 entrevistados tem o grau de escolaridade do ensino fundamental incompleto. A formação superior e pós-graduação representam uma pequena parcela, indicando que a maioria da população tem níveis educacionais mais baixos.

Entrevistados

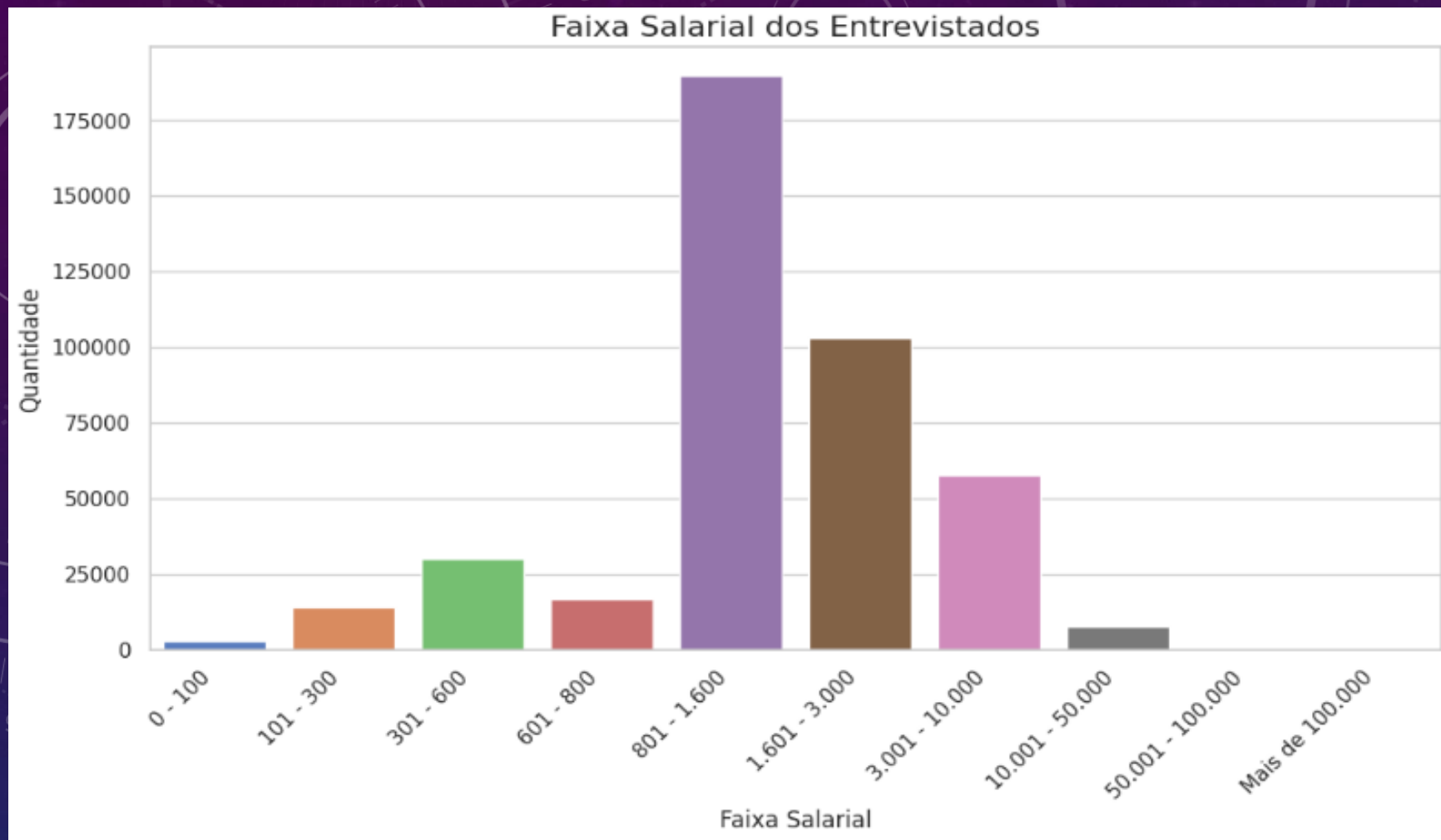


31.140 dos entrevistados tem cargos e ocupações profissionais no setor braçal, a maior parte dessa população necessita aplicar a força de trabalho de forma presencial, o que deixa os profissionais mais expostos ao Covid.



Característica de cargos e ocupações dos entrevistados

Entrevistados

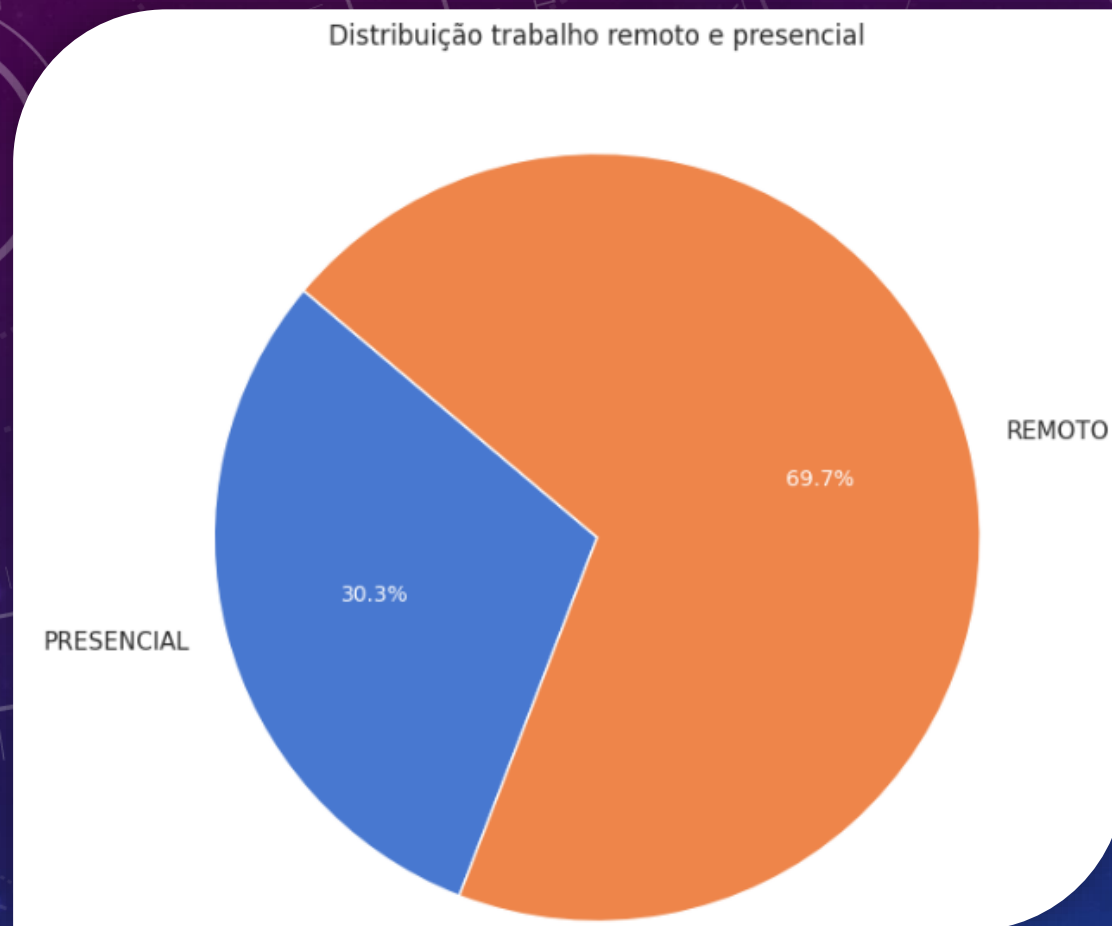


Mais de 175.000
dos
entrevistados
ganham um salário
na faixa de R\$801 a
R\$1.600
Renda que atinge
um salário mínimo



Característica de faixa salarial dos entrevistados

Distribuição do modelo de trabalho

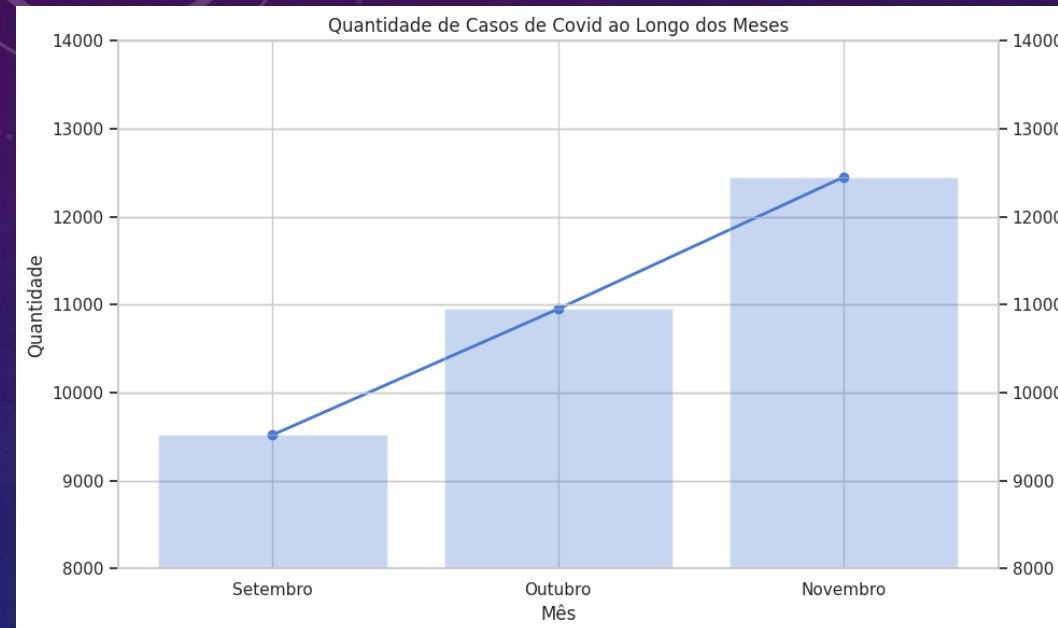
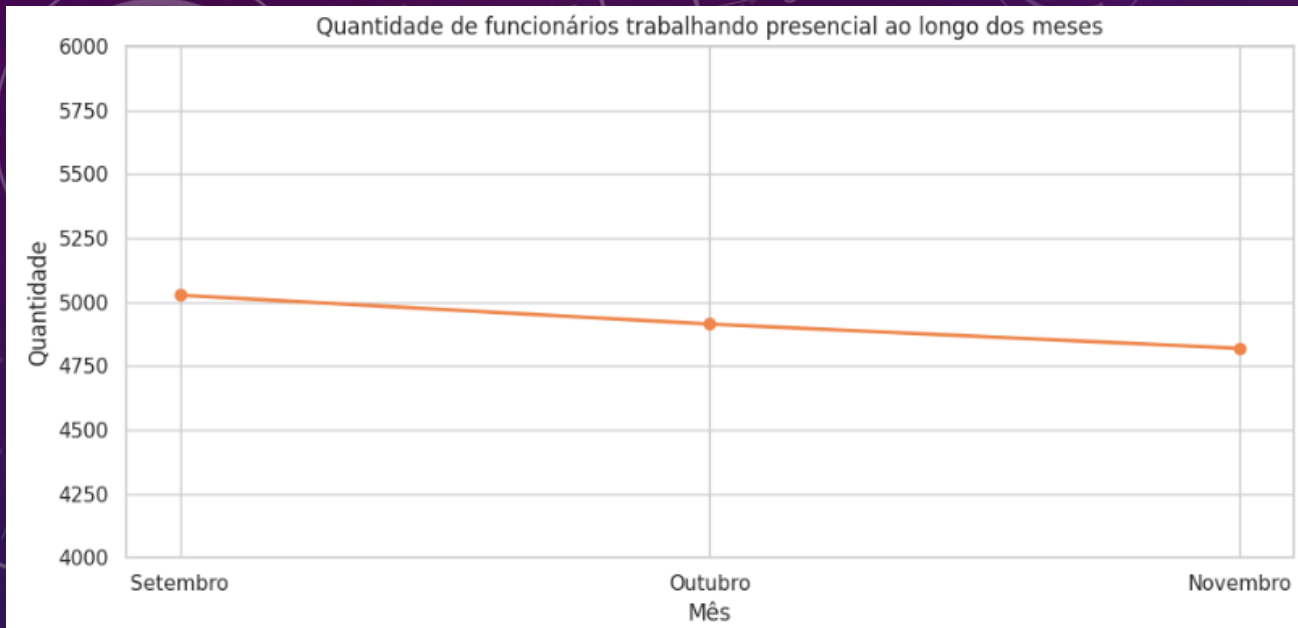


69,7% dos
entrevistados
possuem ocupações
de trabalho remoto



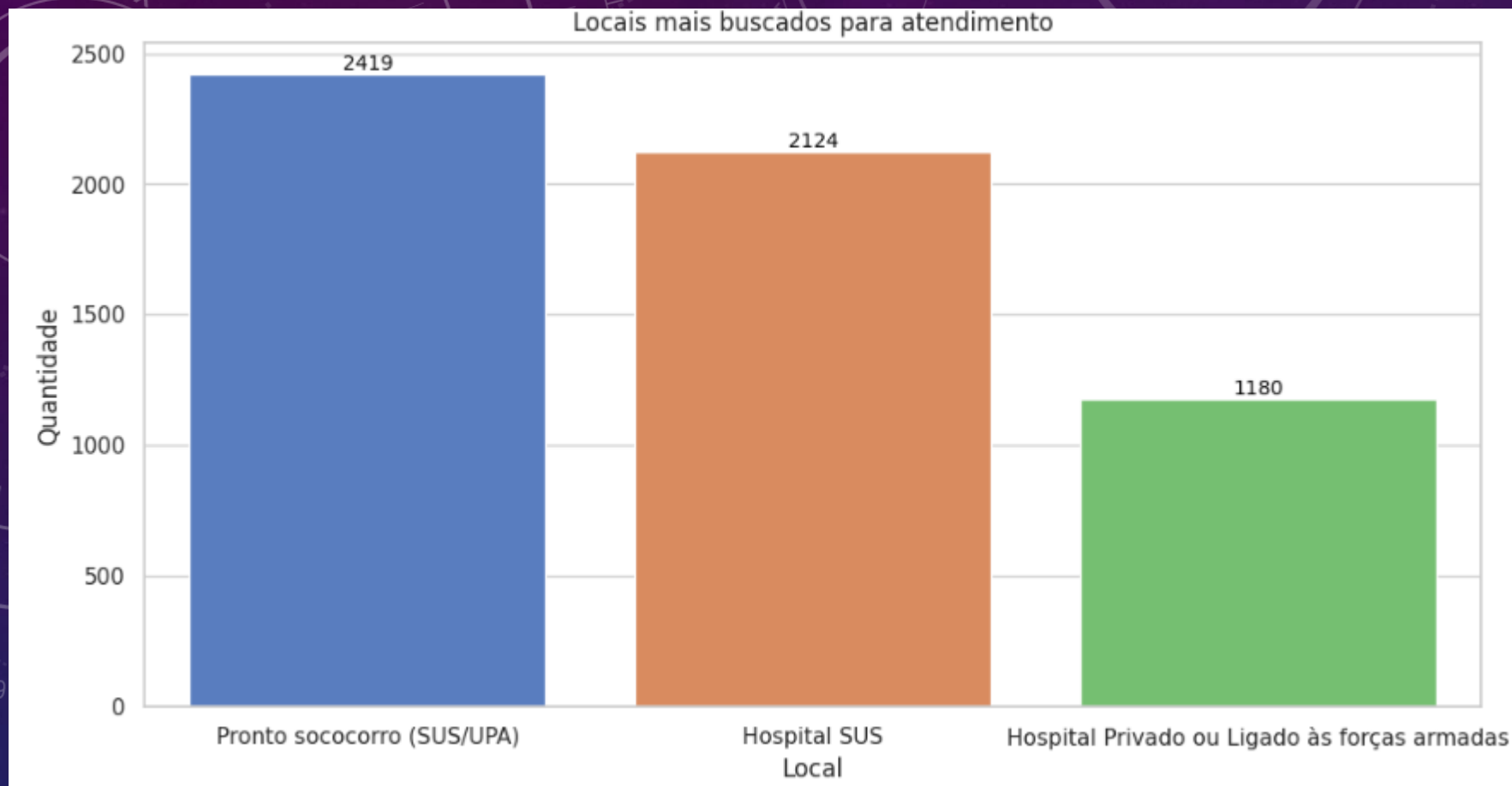
Característica de atuação de modelo de trabalho dos
entrevistados

Entrevistados - Modelo de trabalho presencial



Na análise é possível observar que ao longos dos meses, o modelo de trabalho presencial tem uma leve queda, mesmo com os casos de covid em aumento

Avaliação Médica



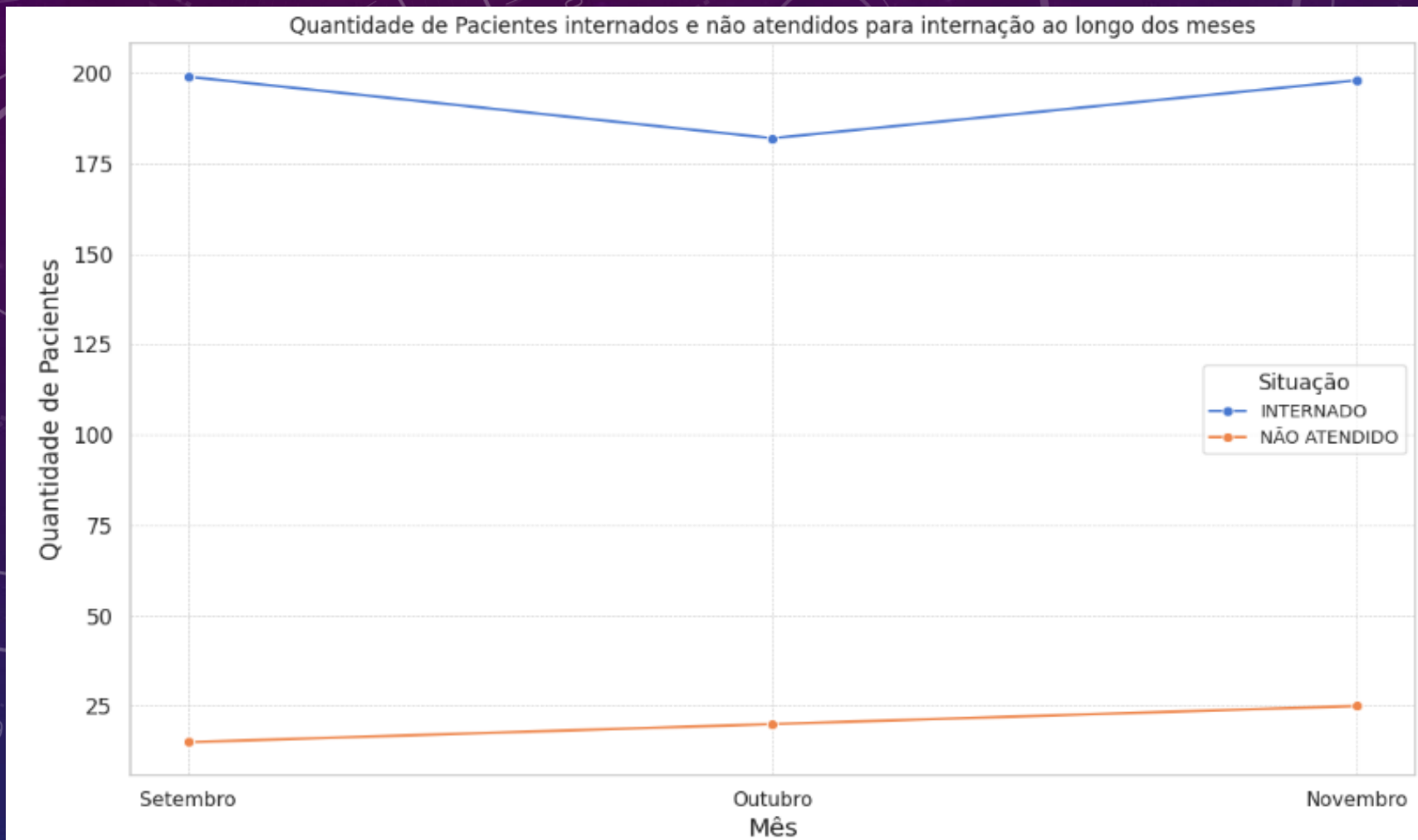
4.543 dos entrevistados buscaram ajuda médica na rede do SUS que foi fundamental para atender grande parte da população.

1.180 entrevistados buscaram ajuda médica na rede de hospitais particulares.



Locais mais buscados pelos entrevistados durante a fase de pandemia da COVID-19

Internações

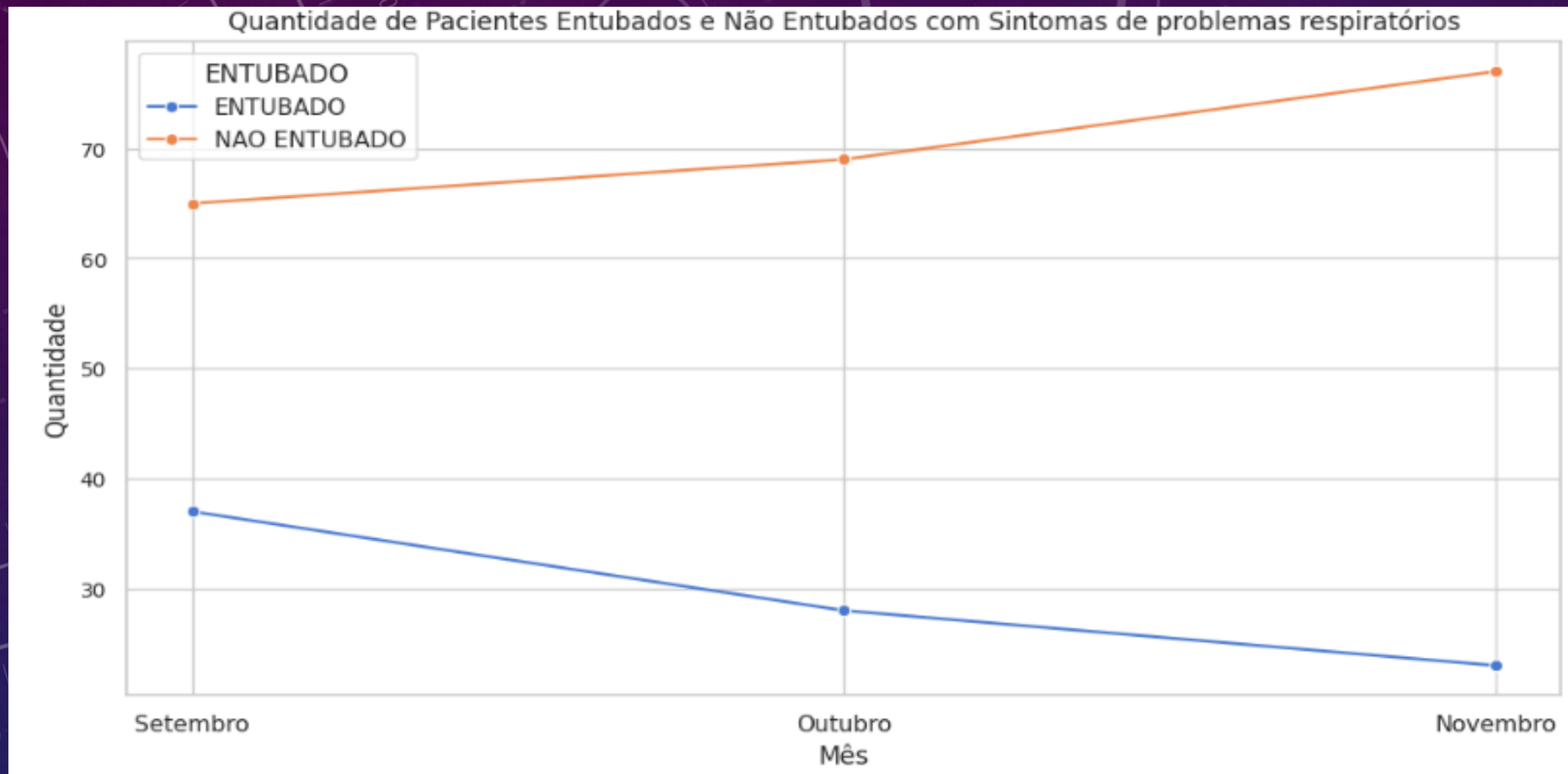


Na análise é possível observar que a quantidade de pacientes não atendidos para internação ao longo dos meses cresceu



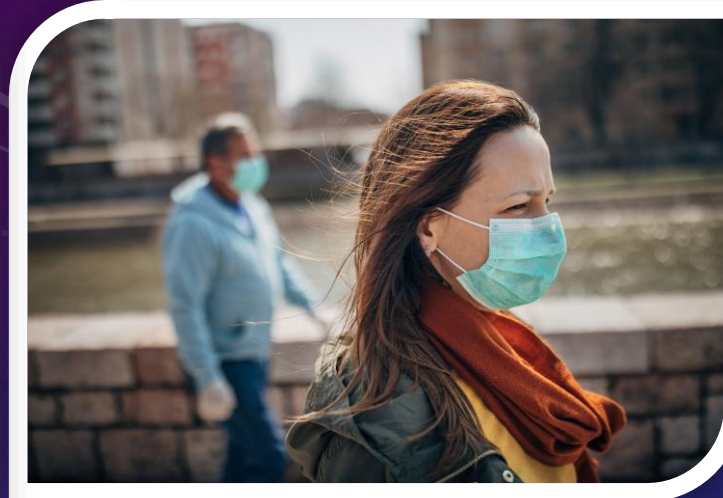
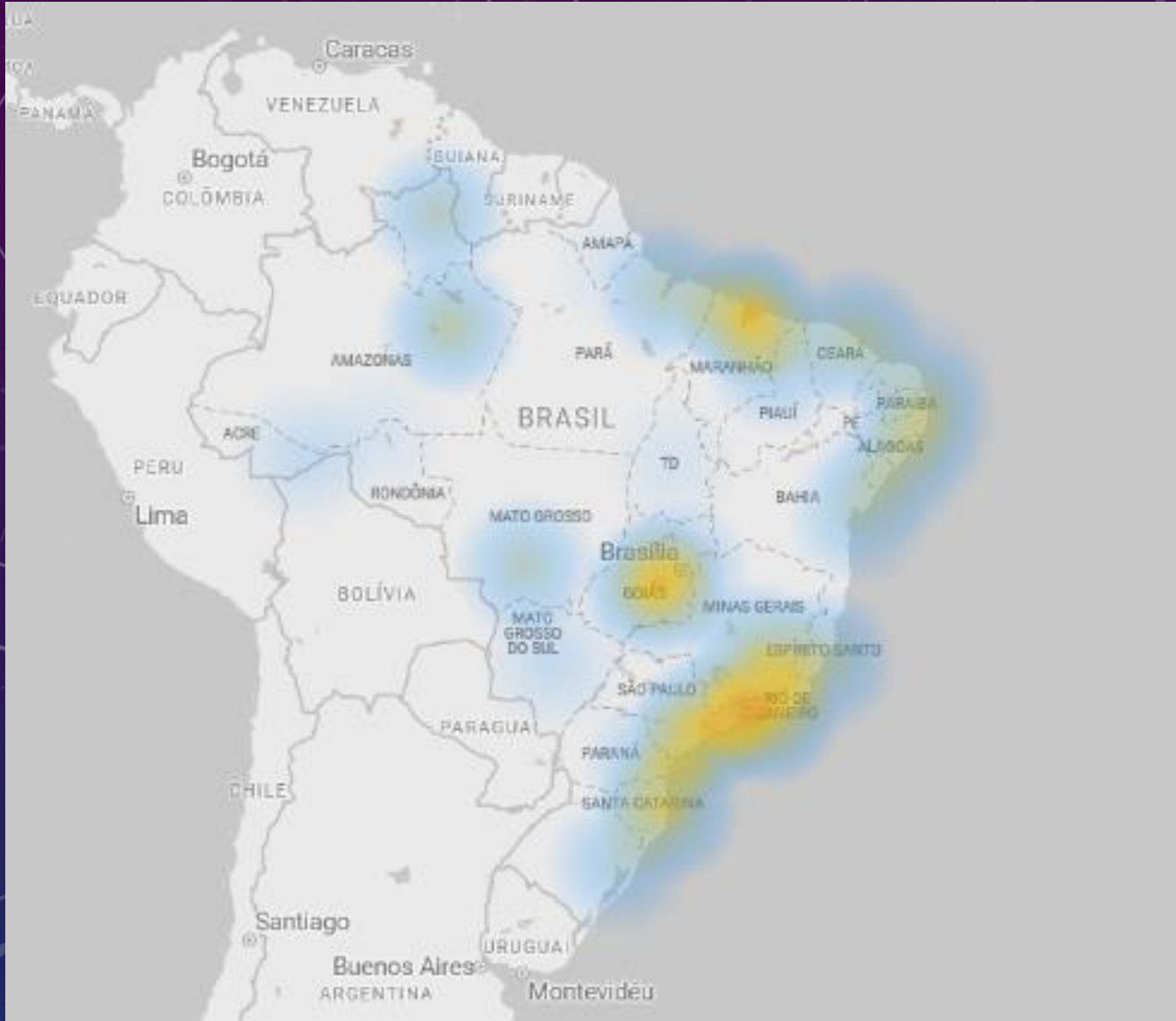
Quantidade de entrevistados internados e não atendidos para internação ao longo dos meses

Pacientes Entubados e não entubados



Na análise a quantidade de pacientes não entubados com sintomas de problemas respiratórios cresceu, essa visão demonstra que os hospitais não estavam preparados para receber uma grande quantidade de pacientes

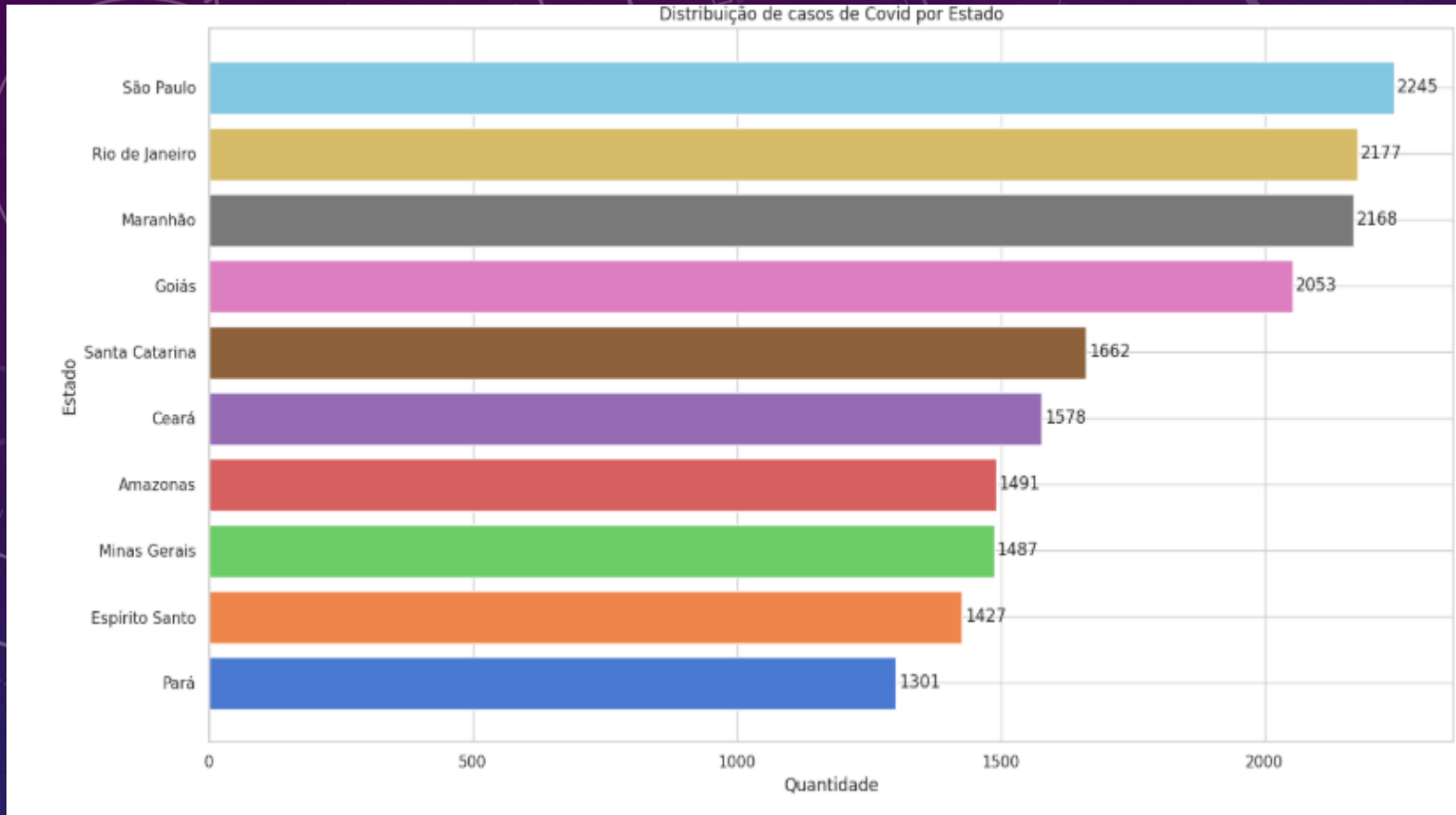
Mapa de infectados no Brasil



Mapa dos estados com mais casos de COVID-19 no Brasil.

**Podemos observar
o crescimento dos
casos de covid nos
estados da região
sudeste**

Mapa de infectados no Brasil



O estado com maior número de infectados é São Paulo. Por ser uma grande cidade populacional o contato entre as pessoas é alto comparado com outros estados.



CONCLUSÃO

Conclusão



- Com a análise realizada na base dos dados disponibilizada, sobre o comportamento da população durante a pandemia de COVID-19 de 2020, podemos observar que o hospital deve usar os dados para a cada dia estar preparado para um surto ou início de uma nova pandemia.
- Durante a análise identificou-se que a grande parte da população utiliza o sistema SUS quando se trata de busca de ajuda médica.
- Outro fator importante demonstra que a maioria da população possui um nível de renda de salário mínimo, e ocupação de trabalho está predominante em trabalhos braçais, com atividades que necessitam da força de trabalho presencial.
- O hospital com base nos dados coletados, devem ficar atento aos sintomas que podem caracterizar o ressurgimento do vírus como Febre, Falta de Paladar ou Olfato e Dificuldade para respirar.
- A pandemia de COVID-19 surpreendeu o mundo e destacou a necessidade de reforçar a infraestrutura hospitalar, garantindo atendimento adequado à população e capacidade para internações em casos graves, especialmente em regiões com maior vulnerabilidade e dependência do SUS.
- Os hospitais devem conscientizar a população para a realização de testagem e restrição pessoal para que o vírus não se espalhe e rapidamente fique fora de controle.
- A infraestrutura dos hospitais devem conseguir identificar os casos positivos e imediatamente isolar esses casos para que o contato com novas pessoas não aconteça e provoque um novo surto. Uma medida que poderia auxiliar na identificação de possíveis casos de Covid, seria um APP para uma triagem digital, com base nos dados anteriores esse APP poderia dizer se o paciente é um possível caso de Covid, o que poderia diminuir o número de visitas nos hospitais.
- Como a busca pelo o sistema do SUS é o mais procurado, os governantes devem realizar parcerias e investimentos para que em caso de uma nova pandemia, a população não fique desprovida de assistência da rede pública

Equipe

Mauro Junior de Oliveira Lima
mauromaurojunior@hotmail.com



Jose Anderson Rodrigues
anderson.j.rodrigues@outlook.com

Samuel Rocha Lima
samuelrrl@gmail.com