

# Análise da PNAD COVID-19: Comportamento da População, Saúde e Impactos Socioeconômicos

## **Tech Challenge – Fase 3**

Nome: Lais Costa Santos Teixeira      RM365830

Curso: Data Analytics

1. Introdução .....	2
2. Fonte de Dados.....	3
3. Engenharia de Dados .....	3
3.1 Arquitetura em Nuvem.....	3
3.2 Organização do Data Lake (S3).....	3
3.3 Amazon Athena .....	5
3.4 AWS Glue .....	6
3.5 Tratamento dos Dicionários de Dados (Notebook – Glue) .....	7
4. Seleção das Variáveis .....	8
4.1 Características Clínicas.....	8
4.2 Comportamento da População .....	8
4.3 Características Econômicas.....	8
5. Modelagem dos Dados .....	8
6. Análises Realizadas.....	9
6.1 Função no Projeto .....	9
6.2 Tipos de Visualizações.....	9
7. Principais Insights e Recomendações para o Hospital .....	10
7.1 Reforço da Atenção Primária e Teleatendimento .....	10
7.2 Foco nos Grupos Mais Afetados.....	11
7.3 Monitoramento dos indicadores chaves .....	11
8. Conclusão .....	13

## 1. Introdução

Este projeto tem como objetivo analisar o comportamento da população brasileira durante a pandemia da COVID-19, com foco em aspectos clínicos, comportamentais e econômicos, utilizando como base os microdados da pesquisa PNAD COVID-19 disponibilizada pelo IBGE.

A análise visa apoiar a tomada de decisão de um hospital, simulando um cenário em que seja necessário planejar ações estratégicas para um eventual novo surto da doença, considerando indicadores relevantes extraídos dos dados.

## 2. Fonte de Dados

Os dados utilizados neste projeto foram obtidos no portal oficial do IBGE:

- PNAD COVID-19
- Microdados mensais da pesquisa
- Dicionário de variáveis
- Questionário aplicado à população

Foram selecionados 3 meses da pesquisa (setembro, outubro e novembro), atendendo ao requisito mínimo estabelecido no projeto.

## 3. Engenharia de Dados

### 3.1 Arquitetura em Nuvem

O projeto foi desenvolvido utilizando serviços da AWS, por meio do ambiente AWS Academy, com as seguintes ferramentas:

- **Amazon S3** – Armazenamento dos dados
- **AWS Glue** – Catálogo de dados e ETL
- **Amazon Athena** – Consulta e análise dos dados via SQL

Essa arquitetura segue o conceito de Data Lake, permitindo escalabilidade e flexibilidade para análises futuras.

### 3.2 Organização do Data Lake (S3)

Os dados foram organizados conforme boas práticas:

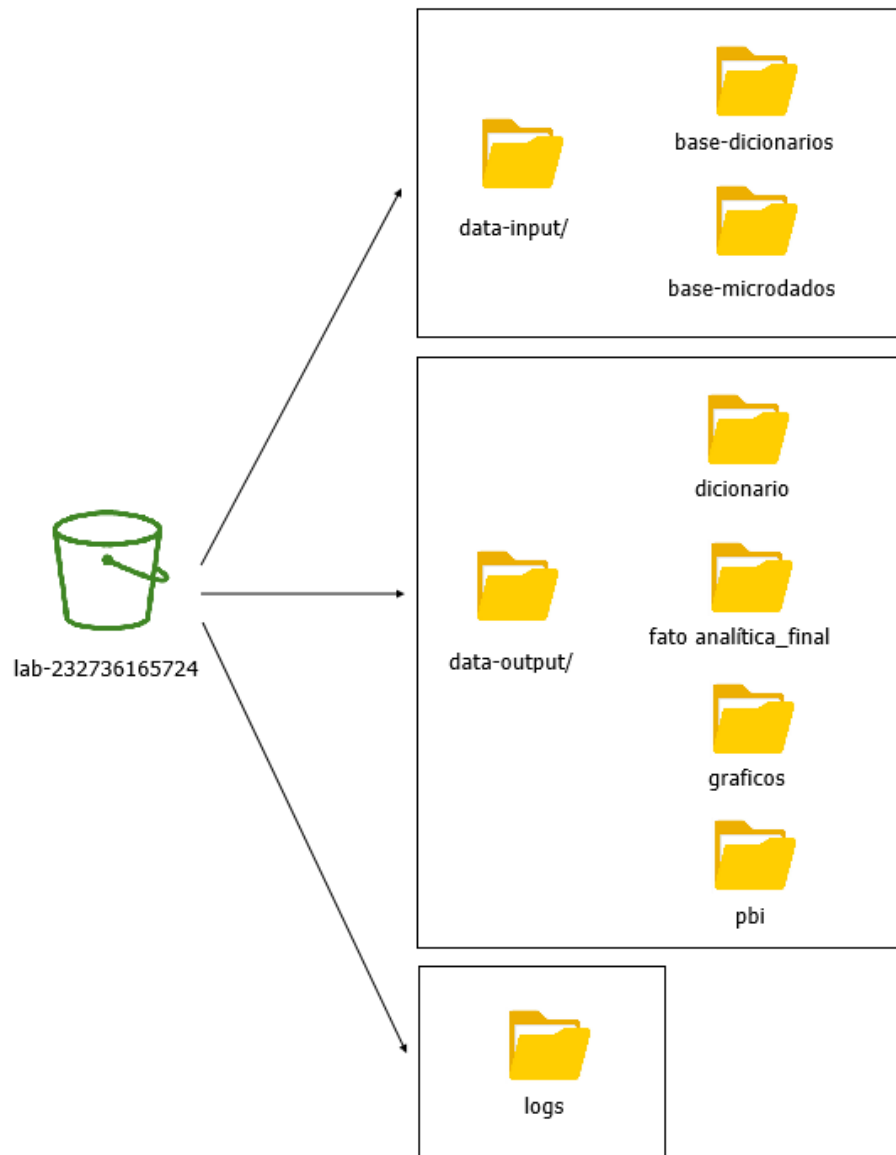


Figura 1: Arquitetura de armazenamento de dados no Amazon S3 (elaboração própria)

lab-232736165724 [Informações](#)

[Objetos](#) [Metadados](#) [Propriedades](#) [Permissões](#) [Métricas](#) [Gerenciamento](#) [Pontos de acesso](#)

**Objetos (4)** [Copiar URI do S3](#) [Copiar URL](#) [Fazer download](#) [Abrir](#) [Excluir](#) [Ações](#) [Criar pasta](#) [Carregar](#)

Os objetos são as entidades fundamentais armazenadas no Amazon S3. Você pode usar o [Inventário do Amazon S3](#) para obter uma lista de todos os objetos em seu bucket. Para outras pessoas acessarem seus objetos, você precisará conceder permissões explicitamente a eles. [Saiba mais](#)

<input type="checkbox"/>	Nome	Tipo	Última modificação	Tamanho	Classe de armazenamento
<input type="checkbox"/>	<a href="#">data-input/</a>	Pasta	-	-	-
<input type="checkbox"/>	<a href="#">data-output_\$folder\$</a>	-	4 Jan 2026 12:07:11 AM -03	0 B	Padrão
<input type="checkbox"/>	<a href="#">data-output/</a>	Pasta	-	-	-
<input type="checkbox"/>	<a href="#">logs/</a>	Pasta	-	-	-

Figura 2: Estrutura de pastas do bucket Amazon S3 do projeto

**data-input/** [Copiar URI do S3](#)

[Objetos](#) [Propriedades](#)

**Objetos (2)** [Copiar URI do S3](#) [Copiar URL](#) [Fazer download](#) [Abrir](#) [Excluir](#) [Ações](#) [Criar pasta](#) [Carregar](#)

Os objetos são as entidades fundamentais armazenadas no Amazon S3. Você pode usar o [Inventário do Amazon S3](#) para obter uma lista de todos os objetos em seu bucket. Para outras pessoas acessarem seus objetos, você precisará conceder permissões explicitamente a eles. [Saiba mais](#)

<input type="checkbox"/>	Nome	Tipo	Última modificação	Tamanho	Classe de armazenamento
<input type="checkbox"/>	<a href="#">base-dicionarios_pnad_covid/</a>	Pasta	-	-	-
<input type="checkbox"/>	<a href="#">base-microdados_pnad_covid/</a>	Pasta	-	-	-

Figura 3: Estrutura de subpastas do bucket Amazon S3 do projeto

**data-output/** [Copiar URI do S3](#)

[Objetos](#) [Propriedades](#)

**Objetos (7)** [Copiar URI do S3](#) [Copiar URL](#) [Fazer download](#) [Abrir](#) [Excluir](#) [Ações](#) [Criar pasta](#) [Carregar](#)

Os objetos são as entidades fundamentais armazenadas no Amazon S3. Você pode usar o [Inventário do Amazon S3](#) para obter uma lista de todos os objetos em seu bucket. Para outras pessoas acessarem seus objetos, você precisará conceder permissões explicitamente a eles. [Saiba mais](#)

<input type="checkbox"/>	Nome	Tipo	Última modificação	Tamanho	Classe de armazenamento
<input type="checkbox"/>	<a href="#">dicionario/</a>	Pasta	-	-	-
<input type="checkbox"/>	<a href="#">fato_analitica_final/</a>	Pasta	-	-	-
<input type="checkbox"/>	<a href="#">fato_analitica_geo/</a>	Pasta	-	-	-
<input type="checkbox"/>	<a href="#">fato_analitica_unico/</a>	Pasta	-	-	-
<input type="checkbox"/>	<a href="#">fato_analitica/</a>	Pasta	-	-	-
<input type="checkbox"/>	<a href="#">graficos/</a>	Pasta	-	-	-
<input type="checkbox"/>	<a href="#">pbi/</a>	Pasta	-	-	-

Figura 4: Estrutura de subpastas do bucket Amazon S3 do projeto

**logs/** [Copiar URI do S3](#)

[Objetos](#) [Propriedades](#)

**Objetos (1)** [Copiar URI do S3](#) [Copiar URL](#) [Fazer download](#) [Abrir](#) [Excluir](#) [Ações](#) [Criar pasta](#) [Carregar](#)

Os objetos são as entidades fundamentais armazenadas no Amazon S3. Você pode usar o [Inventário do Amazon S3](#) para obter uma lista de todos os objetos em seu bucket. Para outras pessoas acessarem seus objetos, você precisará conceder permissões explicitamente a eles. [Saiba mais](#)

<input type="checkbox"/>	Nome	Tipo	Última modificação	Tamanho	Classe de armazenamento
<input type="checkbox"/>	<a href="#">Unsaved/</a>	Pasta	-	-	-

Figura 5: Estrutura de subpastas do bucket Amazon S3 do projeto

### 3.3 Amazon Athena

O Amazon Athena foi utilizado para consulta e validação dos dados armazenados no Amazon S3.

As principais atividades realizadas no Athena foram:

- **Criação do banco de dados:** Workspace;
- **Execução de consultas SQL** para exploração inicial dos dados (análise exploratória).

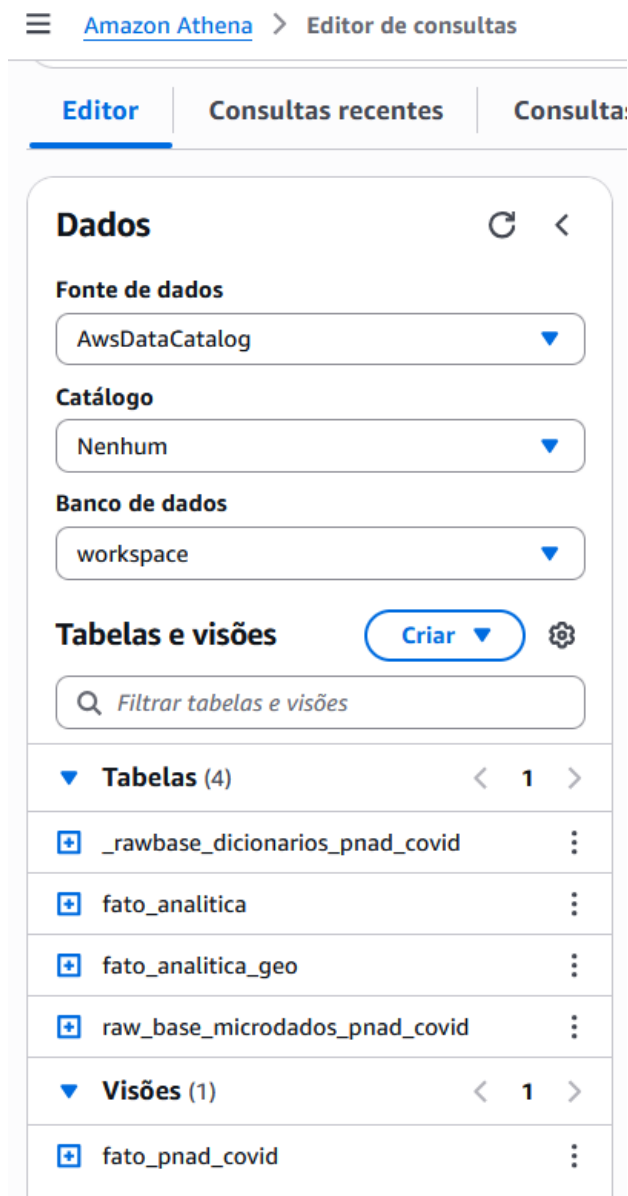


Figura 6: Banco de dados, tabelas e views criadas no Amazon Athenas.

### 3.4 AWS Glue

O AWS Glue foi utilizado como ferramenta de catálogo e preparação dos dados armazenados no Amazon S3.

As principais atividades realizadas no Glue foram:

- **Criação do Data Catalog**, permitindo o registro das tabelas a partir dos arquivos armazenados no S3.
- **Definição do esquema dos dados**, identificando colunas, tipos de dados e estrutura dos arquivos.
- **Utilização de crawlers** para automatizar o processo de descoberta, catalogação e atualização dos dados armazenados no Amazon S3.
- Organização das tabelas de forma padronizada, facilitando a integração com ferramentas analíticas como o Amazon Athena e o PySpark (Notebooks)

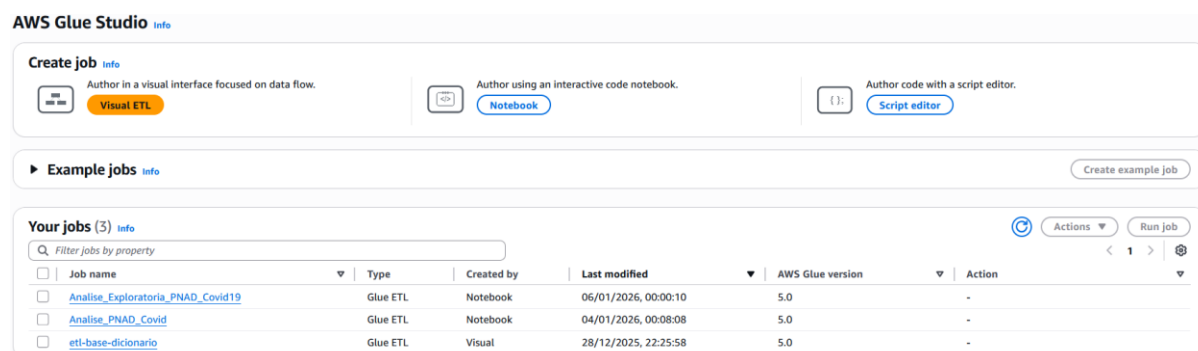


Figura 7: Jobs criados para tratamento dos dados na AWS Glue Studio, utilizando notebooks.

### 3.5 Tratamento dos Dicionários de Dados (Notebook – Glue)

Os dicionários da PNAD COVID-19 apresentam informações com:

- células mescladas
- descrições repetidas
- códigos e categorias em linhas diferentes

Para facilitar o uso analítico, foi realizada a normalização dos dicionários, transformando-os em tabelas estruturadas contendo:

- Código da variável
- Descrição da variável
- Código da categoria
- Descrição da categoria

Esses dicionários foram tratados e utilizados como **tabelas dimensão**, permitindo a correta interpretação dos códigos presentes nos microdados.

## 4. Seleção das Variáveis

Foram selecionadas 20 variáveis, conforme exigido no projeto, distribuídas nos três eixos obrigatórios:

### 4.1 Características Clínicas

- Sintomas relatados (febre, tosse, dor de garganta, etc.)
- Dificuldade respiratória

### 4.2 Comportamento da População

- Procura por atendimento médico
- Providências tomadas para tratar os sintomas
- Se houve Internações

### 4.3 Características Econômicas

Situação ocupacional

- Afastamentos Temporários
- Período do afastamento
- Motivo do afastamento

A escolha dessas variáveis foi realizada com foco na geração de **indicadores relevantes para o planejamento hospitalar**.

## 5. Modelagem dos Dados

A modelagem seguiu um modelo analítico simplificado, próximo ao **modelo estrela**, contendo:

- **Tabela Fato:** registros mensais da PNAD COVID-19
- **Tabelas Dimensão:** Perguntas, Resposta, UF, Capital, Sexo, Raça, Escolaridade.



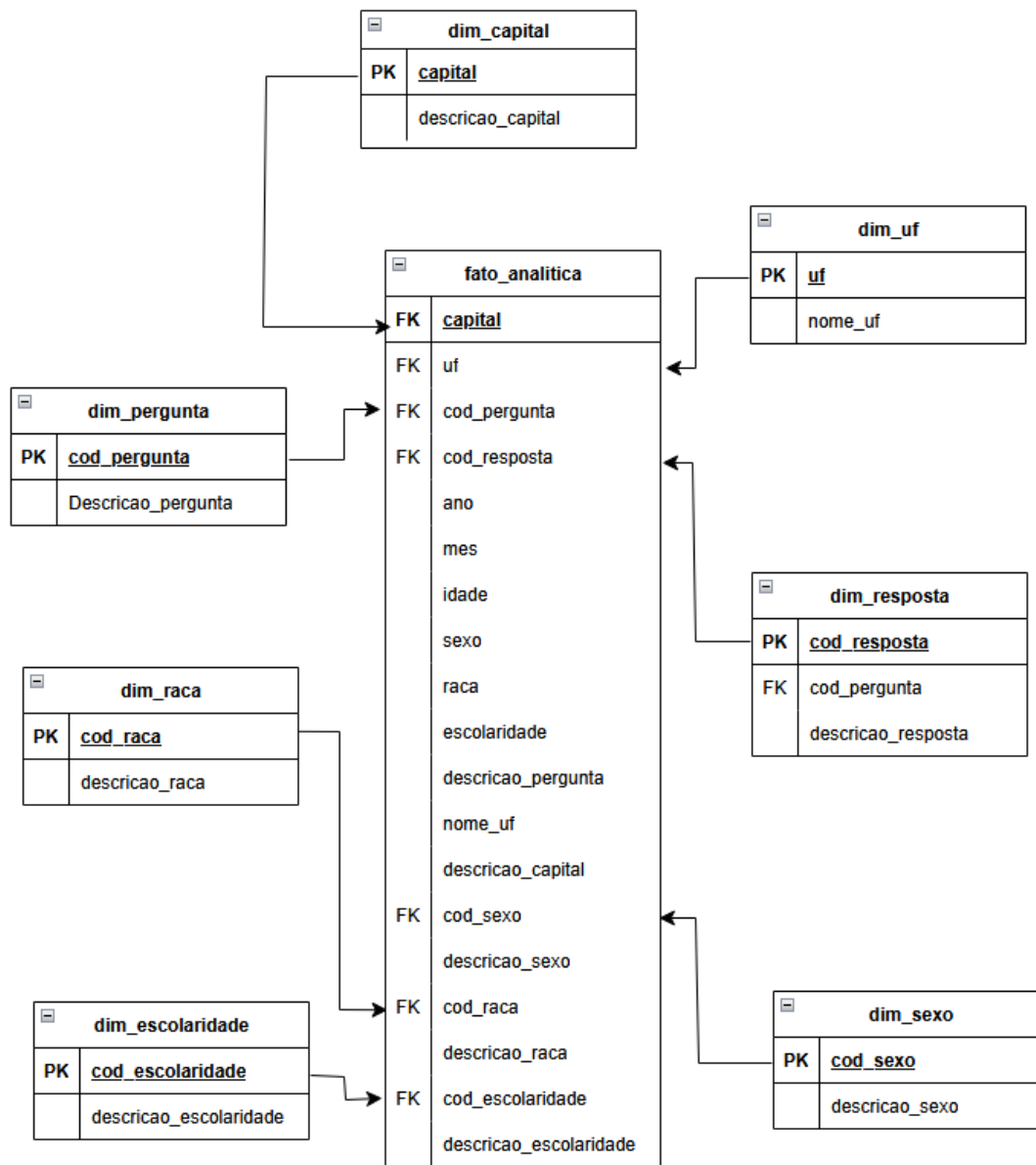


Figura 8: Esquema estrela

## 6. Análises Realizadas

O Power BI foi utilizado como ferramenta de Data Visualization e Storytelling.

### 6.1 Função no Projeto

- Conectar-se aos arquivos CSV gerados pelo Glue;
- Criar dashboards interativos;
- Facilitar a interpretação dos dados por meio de gráficos e indicadores;
- Apoiar a análise exploratória e a comunicação dos resultados.

## 6.2 Tipos de Visualizações

Foram usados principalmente gráficos de barras e de colunas divididos em 3 páginas, uma para cada tópico (Sintomas, Comportamento e Econômico).

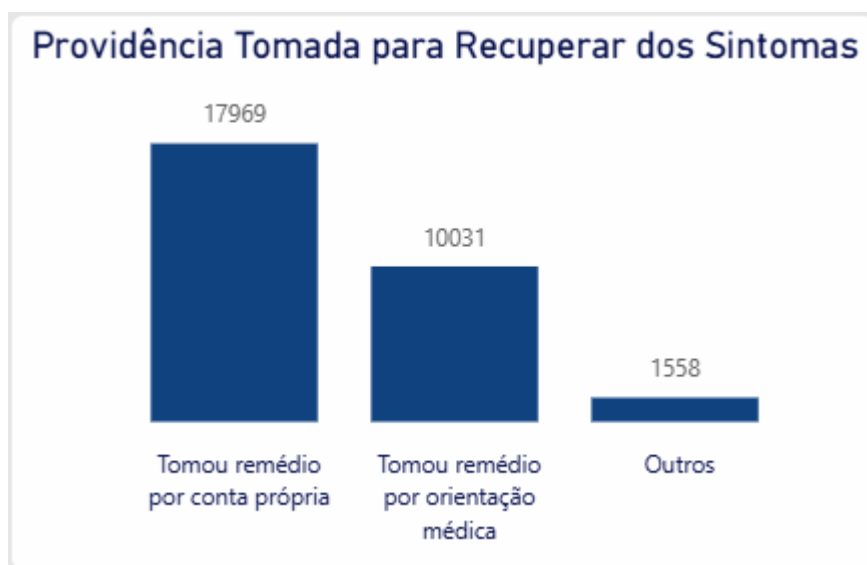
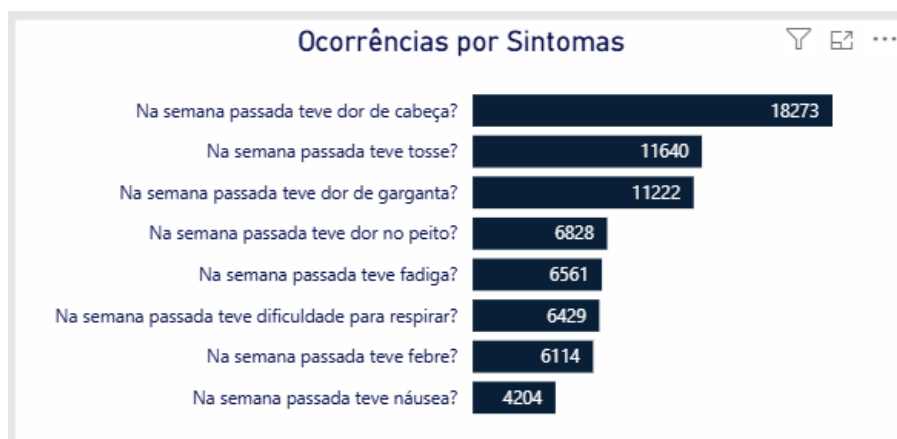
Além das bases geradas pela Glue, também houve a necessidade de criação de medidas para facilitar alguns cálculos de comparação entre os tópicos.

**Observação:** O arquivo pdf será anexado ao final da documentação e o arquivo pbix será anexado a pasta do projeto.

## 7. Principais Insights e Recomendações para o Hospital

### 7.1 Reforço da Atenção Primária e Teleatendimento

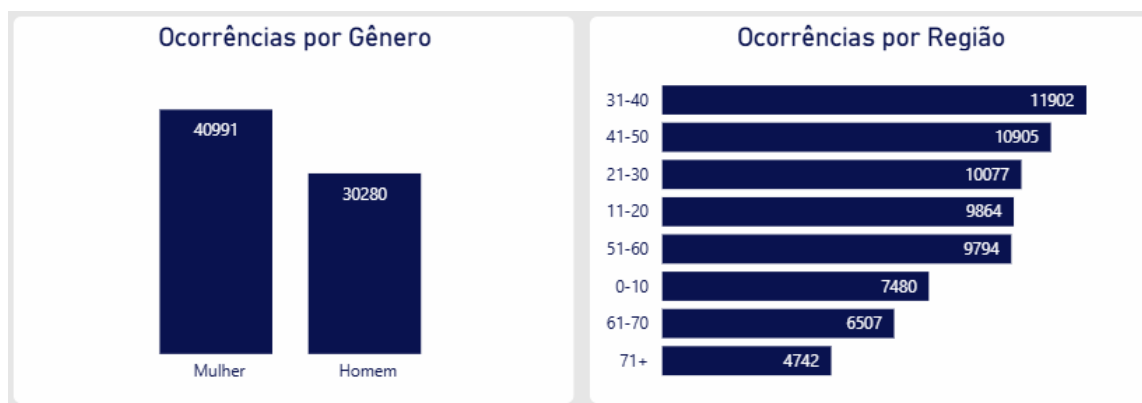
- Alta ocorrência de sintomas leves e moderados.
- Grande parte da população buscou orientação ou tratamento fora do hospital.



Para esses casos, a recomendação é expandir os canais de telemedicina e orientação remota, principalmente para a triagem de sintomas mais leves. Dessa forma, o Hospital assegura que mesmo aqueles que não desejam sair de casa recebam uma orientação médica adequada para o tratamento dos sintomas.

## 7.2 Foco nos Grupos Mais Afetados

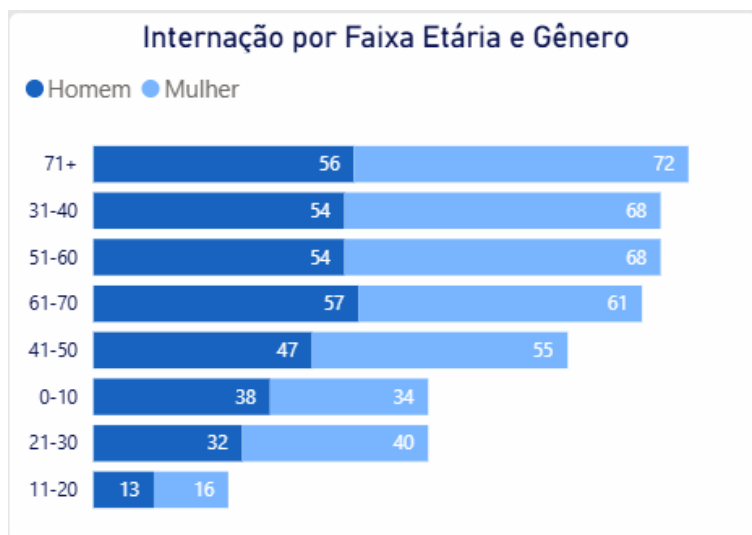
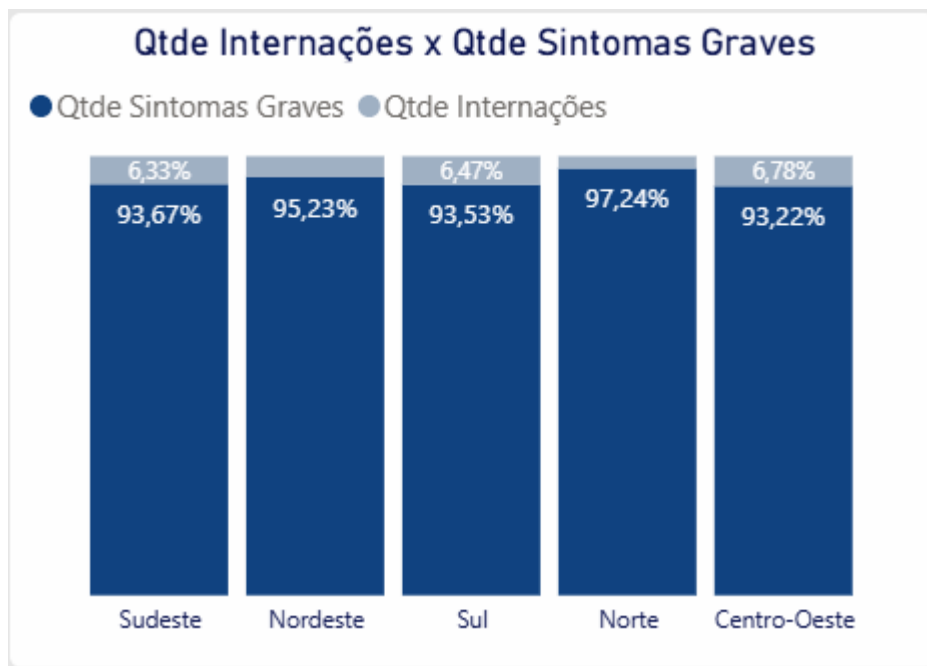
- Mulheres e pessoas entre 31 e 60 anos apresentaram maior ocorrência de sintomas e afastamentos.
- Maior impacto em regiões Sudeste e Nordeste.



O Hospital pode criar campanhas de prevenção e acompanhamento para esses grupos. E priorizar a conscientização dos riscos da automedicação.

## 7.3 Monitoramento dos indicadores chaves

- Diferenças relevantes por região, sexo e faixa etária.



Manter e acompanhar indicadores como região, sintomas graves, perfil do público e procura por atendimento traz diversas vantagens estratégicas para o Hospital:

- **Visão territorial clara:** permite identificar quais regiões apresentam maior demanda ou incidência de sintomas, direcionando recursos e equipes de forma mais eficiente.
- **Conhecimento do público atendido:** possibilita entender melhor o perfil dos pacientes (faixa etária, condições pré-existent, hábitos), o que ajuda na personalização das orientações médicas.
- **Gestão da procura por atendimento:** auxilia na previsão de picos de demanda, evitando sobrecarga de serviços.

Com dados atualizados, a instituição pode agir de forma proativa, antecipando medidas de prevenção, campanhas de conscientização ou reforço de equipes.

## **8. Conclusão**

O uso dos dados da PNAD COVID-19 permitiu compreender aspectos fundamentais do comportamento da população durante a pandemia, fornecendo subsídios importantes para o planejamento hospitalar em cenários futuros.

A arquitetura em nuvem adotada possibilita escalabilidade, reutilização da base e aprofundamento das análises, reforçando a importância da engenharia de dados aliada à análise estratégica.

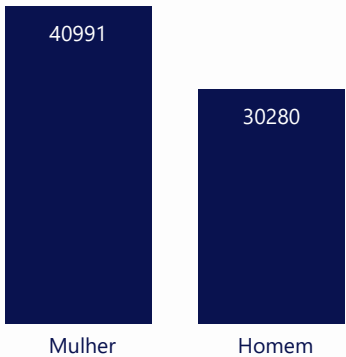
# Hospital Geral

## Sintomas

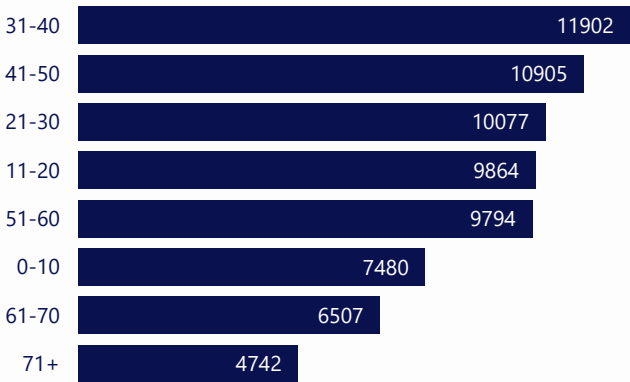
Ocorrências por Sintomas



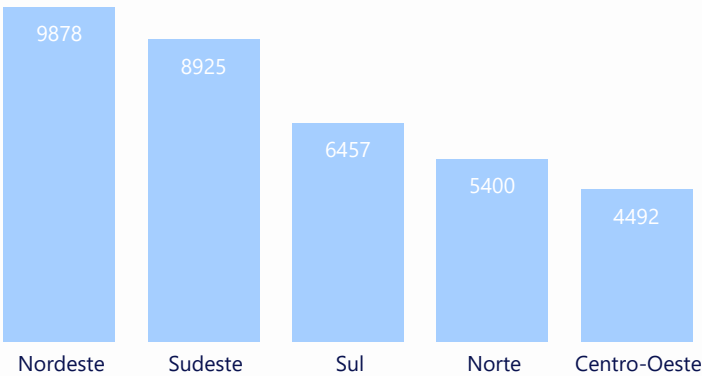
Ocorrências por Gênero



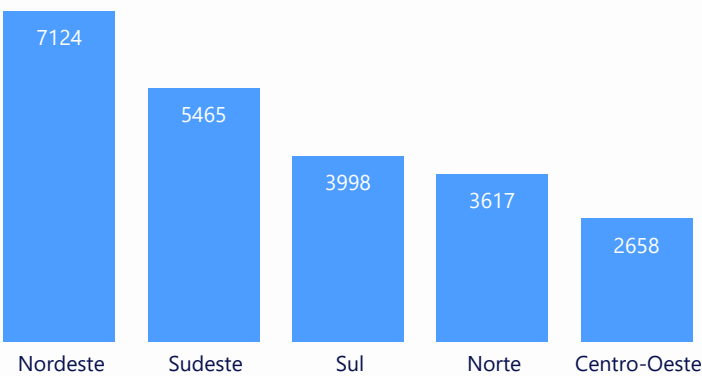
Ocorrências por Região



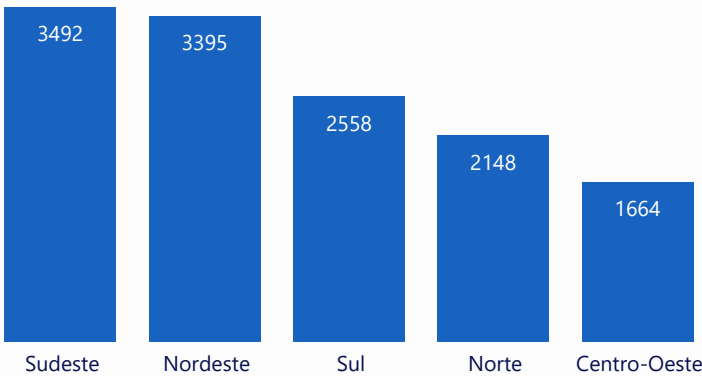
Sintomas Geral



Respiratório Leve



Respiratório Grave

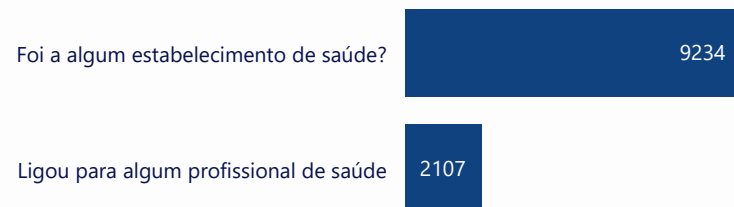


**Análise:** Nos meses de setembro, outubro e novembro de 2020, o sintoma mais frequentemente relatado foi dor de cabeça, indicando predominância de quadros leves na população analisada. A maior ocorrência de sintomas foi observada entre mulheres e na faixa etária de 31 a 40 anos, grupo majoritariamente economicamente ativo. As regiões Nordeste e Sudeste concentraram o maior volume de sintomas gerais e respiratórios leves, enquanto o Sudeste apresentou maior incidência de sintomas respiratórios graves, sugerindo maior pressão sobre a rede de saúde regional.

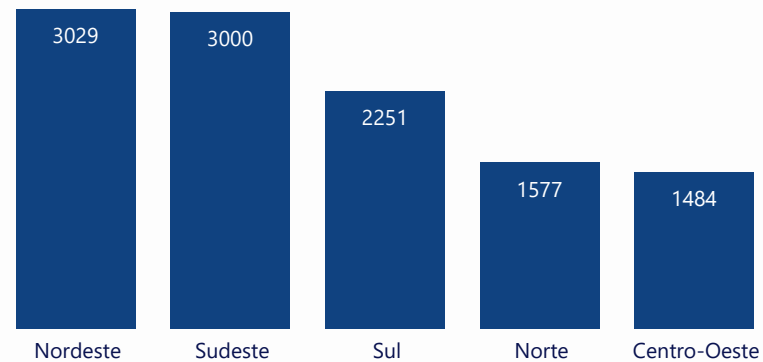
# Hospital Geral

## Comportamento

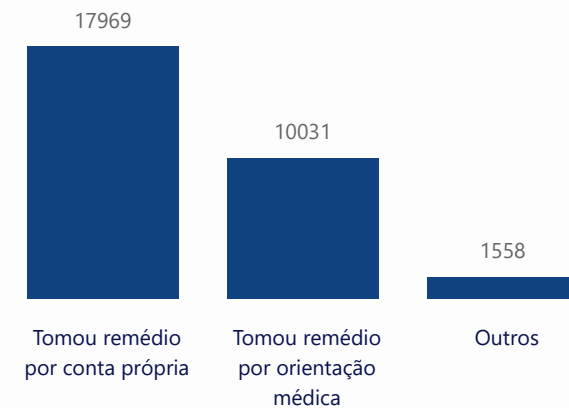
### Respondentes que Procuraram Atendimento



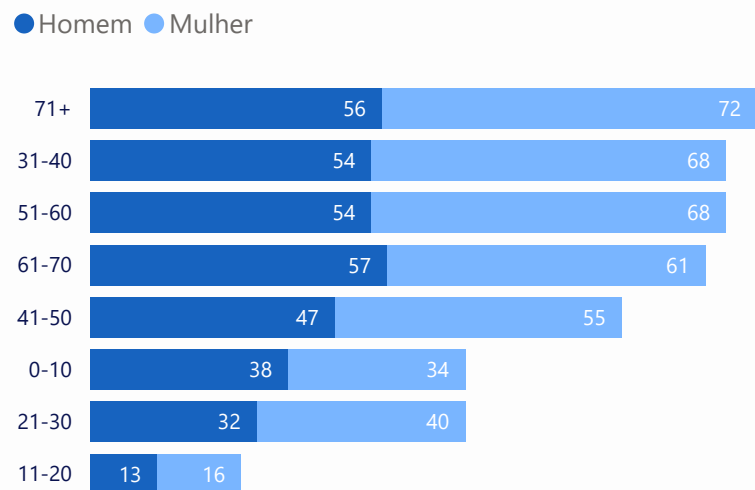
### Respondentes que Procuraram Atendimento por Região



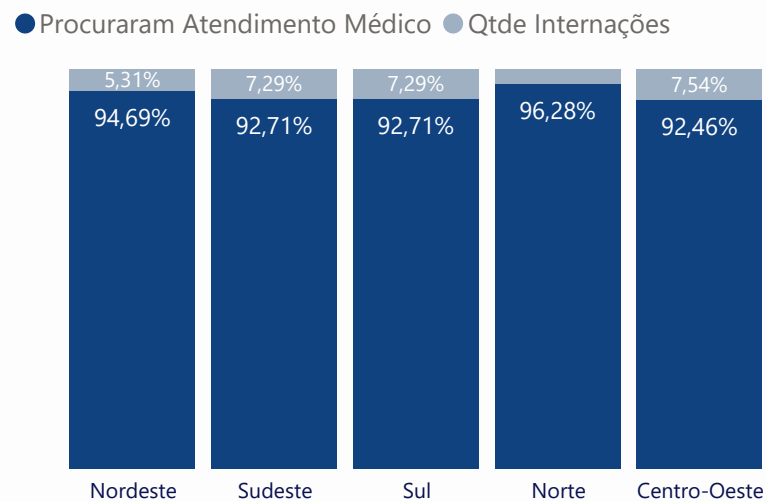
### Providência Tomada para Recuperar dos Sintomas



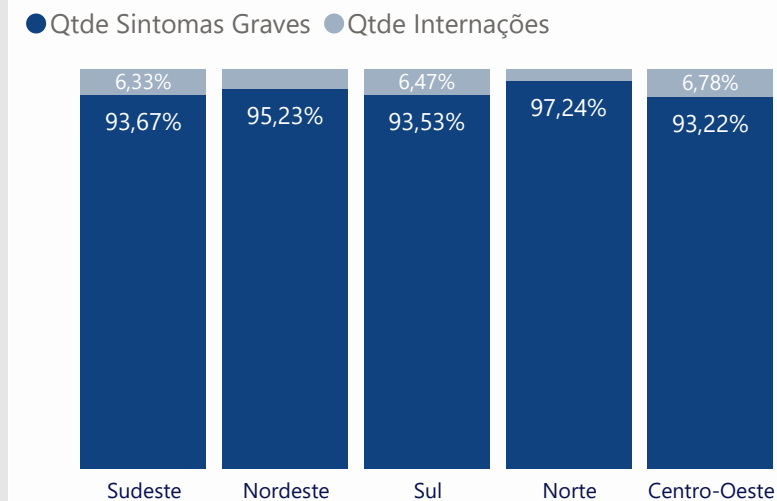
### Internação por Faixa Etária e Gênero



### Procura por Atendimento x Internação



### Qtde Internações x Qtde Sintomas Graves



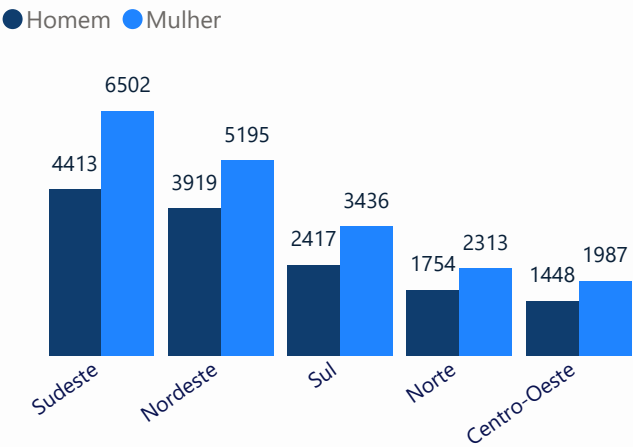
**Análise:** Nos três meses analisados, cerca de 11,3 mil pessoas buscaram atendimento médico, enquanto aproximadamente 18 mil trataram os sintomas por conta própria. As internações concentraram-se no Sudeste, sendo a maioria idosos, e representaram menos de 8% dos atendimentos médicos e menos de 7% dos casos com sintomas respiratórios graves.

Fonte: PNAD-COVID-19 do IBGE (<https://covid19.ibge.gov.br/pnad-covid/>)

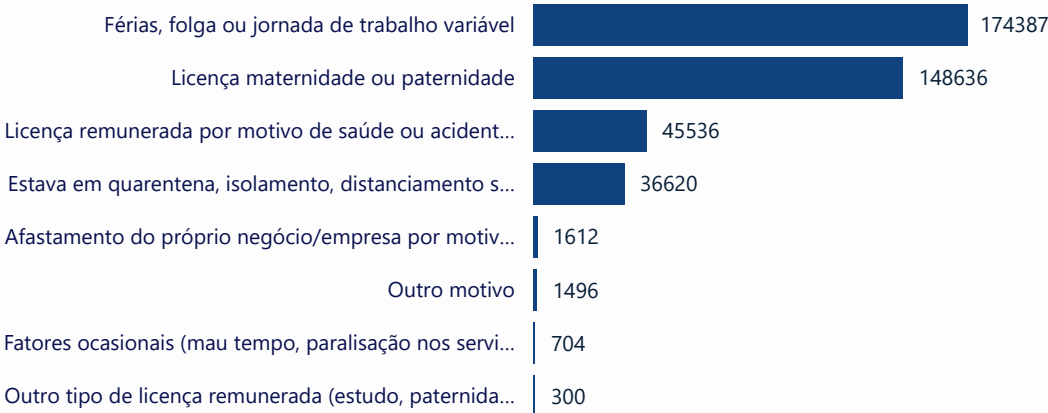
# Hospital Geral

## Econômico

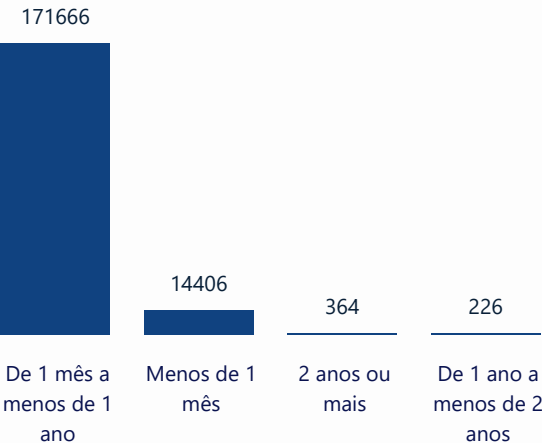
Qtde Afastamentos Temporários



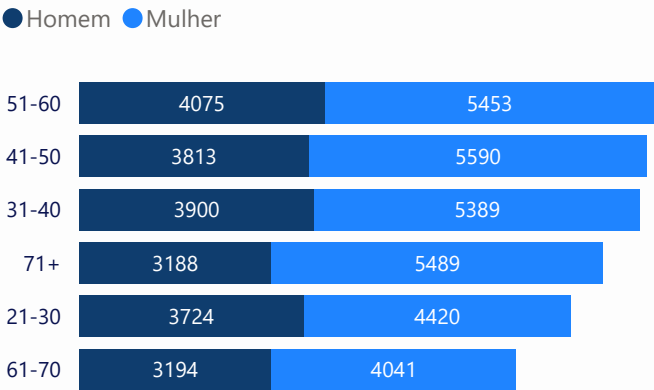
Principal Motivo do Afastamento



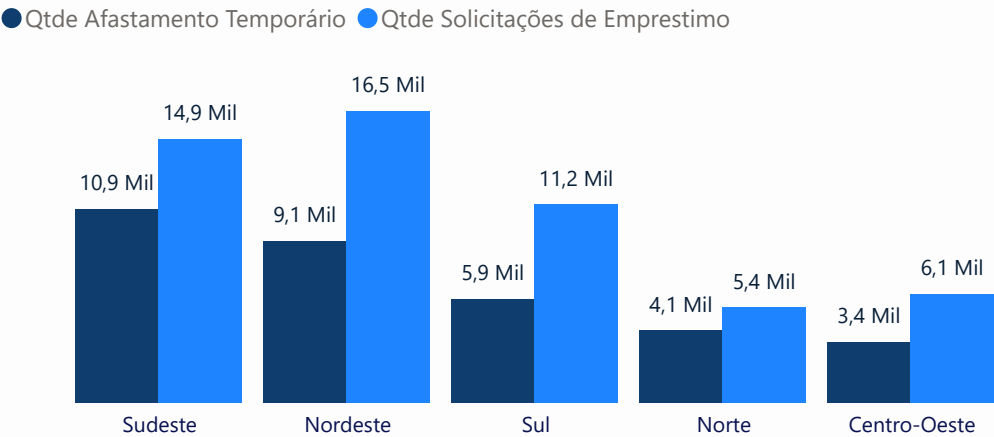
Tempo de Afastamento



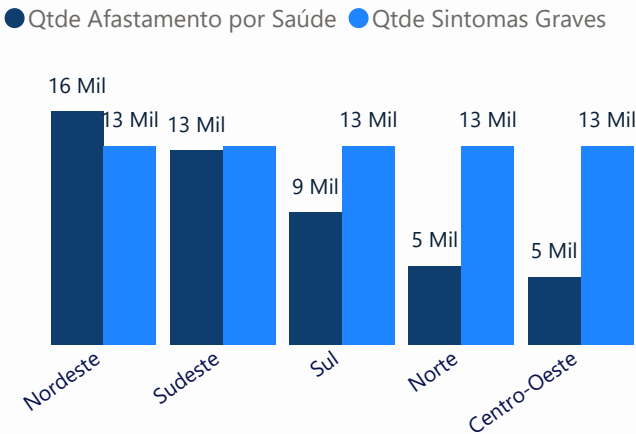
Afastamento por Saúde ou Distanciamento Social



Qtde Afastamento Temporário x Qtde Solicitações de Empréstimo



Afastamento por Saúde x Sintomas Graves



**Análise:** As mulheres concentraram o maior número de afastamentos, especialmente nas regiões Sudeste e Nordeste. O principal motivo foi férias e/ou folga. O distanciamento social, ficou com mais de 36 mil ocorrências. A maioria dos afastamentos durou entre 1 mês e menos de 1 ano, sendo os afastamentos por saúde mais frequentes entre mulheres de 41 a 60 anos. A solicitação de empréstimos superou o número de afastamentos, indicando maior vulnerabilidade econômica. Nas regiões Nordeste e Sudeste, os afastamentos por saúde acompanham os casos graves, enquanto nas demais regiões os sintomas graves superam os afastamentos.