



## EDA Pandemia

### 1. Visão Geral

#### 1.1. Quantidade de Linhas e Colunas

Após a limpeza e o tratamento dos dados brutos do DATASUS, o *dataframe* final resultou em **241.285 entradas**, cada uma composta por **23 variáveis** distintas para análise.

#### 1.2. Colunas do Dataset

O conjunto de dados consolidado para o período de pandemia do COVID-19 contém **241.285 linhas e 23 colunas**, divididas nas seguintes categorias de informações:

- **Geografia e Localização:** UF\_ZI (Unidade Federativa), MUNIC\_RES (Município de Residência) e MUNIC\_MOV (Município de Movimentação/Hospital).
- **Demografia do Paciente:** NASC (Data de Nascimento), IDADE, SEXO e RACA\_COR.
- **Temporalidade:** ANO\_CMPT (Ano de competência), MES\_CMPT (Mês de competência), DT\_INTER (Data de Internação) e DT\_SAIDA (Data de Saída).
- **Informações Clínicas e Diagnósticos:** DIAG\_PRINC (Causa principal), DIAG\_SECUN, CID\_ASSO (CID Associado), CID\_MORTE e ESPEC (Especialidade).
- **Gestão e Internação:** CAR\_INT (Caráter da internação), IDENT, COBRANCA, COMPLEX (Complexidade), QT\_DIARIAS (Quantidade de diárias) e DIAS\_PERM (Dias de permanência).
- **Desfecho:** MORTE (Indicador binário de óbito).

#### 1.3. Tipo dos dados

#### 1.4. Valores nulos



# Documentação Análise Exploratória de Dados

**Consistência da Base:** Foi feito um tratamento de dados no dataset inicial adquirido no Datasus, assim o dataset final não apresenta valores nulos (missing values) em nenhuma das 23 colunas analisadas.

## 1.5. Valores Duplicados

Para garantir a fidedignidade dos indicadores, a base de dados passou por um processo de limpeza e desduplicação antes das análises:

**Volume Bruto:** o conjunto de dados inicial continha **241.285 registros**.

Identificação de duplicatas: foram identificados **2653** registros duplicados (aproximadamente **1,1%** da base total).

**Justificativa:** essas duplicatas decorrem de processamentos administrativos e correções de faturamento no sistema do SIH-SUS .

**Tratamento:** foi utilizado uma chave de internação para classificar o evento de uma internação como mostrado a seguir

```
chave_internacao = [ 'UF_ZI', 'ANO_CMPT', 'MUNIC_RES', 'DT_INTER',  
'DT_SAIDA', 'DIAG_PRINC', 'SEXO', 'IDADE', 'MORTE' ]
```

**Universo de Estudo Final:** após a limpeza, a análise foi conduzida sobre 238.632 internações únicas, assegurando que cada evento assistencial fosse contabilizado apenas uma vez.

## 1.6. Recorte Temporal

O período analisado foi de janeiro de 2020 a dezembro de 2022, período que compõe a pandemia do COVID-19.

# 2. Panorama Geral

## 2.1. Número total de internações

Após o processo de limpeza e desduplicação, o estudo consolidou um universo de **238.632 internações únicas** no estado de São Paulo entre 2020 e 2022.

## 2.2. Média Mensal de Internações

O número médio de internações por mês no período analisado foi de aproximadamente **6.628 internações**, calculado a partir da contagem mensal de internações únicas.

## 2.3. Quantidade de municípios residente e de internação



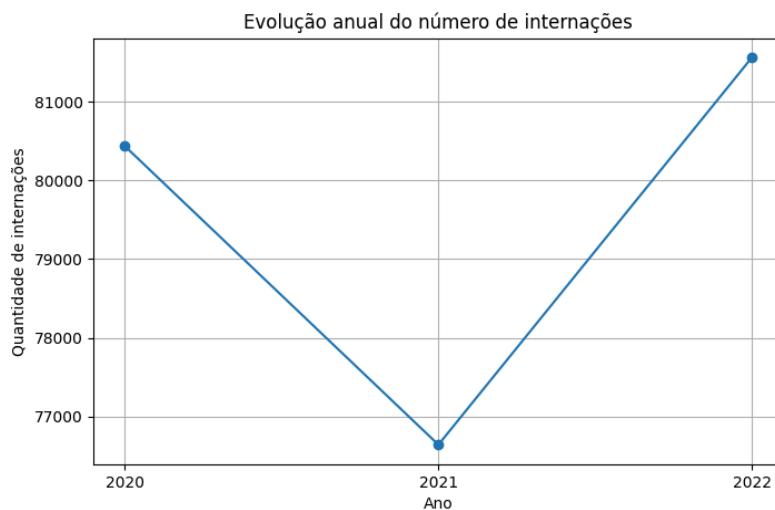
# Documentação Análise Exploratória de Dados

Enquanto os pacientes são provenientes de **843 municípios** diferentes, a rede hospitalar está concentrada em apenas **299 cidades**. Essa disparidade revela uma centralização dos serviços de saúde mental, obrigando a maioria dos pacientes a se deslocarem de suas cidades de origem para obter assistência hospitalar, evidenciando uma desigualdade no acesso à saúde.

## 3. Análise Temporal

### 3.1. Evolução Anual das Internações

Observa-se uma queda acentuada em 2021, possivelmente refletindo o represamento de demandas durante a crise sanitária, com uma retomada expressiva em 2022, quando o volume de atendimentos voltou a subir e superou o patamar inicial do período pandêmico, conforme mostra o gráfico:



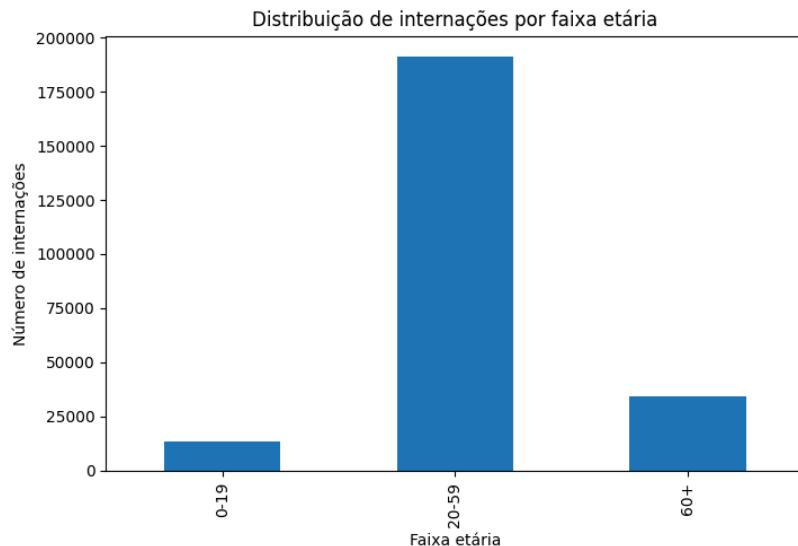
## 4. Perfil Sociodemográfico

### 4.1. Distribuição de Idade

O volume de internações por saúde mental é predominantemente concentrado na faixa de **20–59 anos**, que supera em mais de cinco vezes o total de idosos (60+). Destaca-se a vida adulta como o período de maior vulnerabilidade ou registro de crises agudas, não mudando muito em relação ao período antes da pandemia.

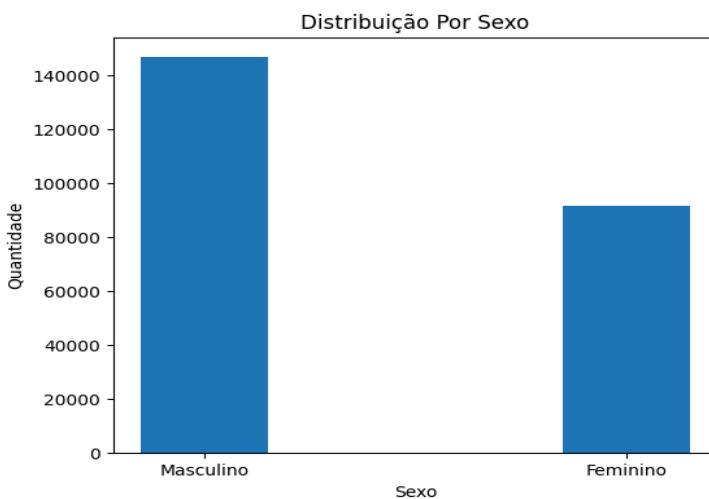


## Documentação Análise Exploratória de Dados



### 4.2. Distribuição de Sexo

O público masculino representa a maioria expressiva das internações (**61,60%**). Esse dado indica uma maior prevalência de crises agudas ou menor busca por tratamento preventivo entre os homens, resultando em mais hospitalizações. Uma das hipóteses é o preconceito que homens sofrem ao falar sobre saúde mental.



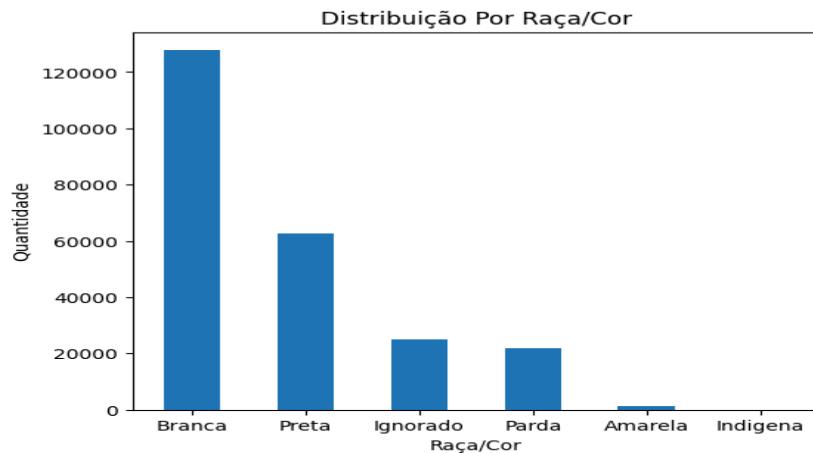
### 4.3. Distribuição de Raça/Cor

A predominância de **pacientes brancos (53%)** contrasta com a soma de **pretos e pardos (35%)**, sugerindo disparidades no acesso ou no perfil das internações. Chama a atenção o alto volume de dados **ignorados (10%)**, o que prejudica a precisão das políticas de equidade, enquanto a baixa presença de



# Documentação Análise Exploratória de Dados

indígenas e amarelos aponta para uma possível exclusão ou subnotificação severa desses grupos no sistema.



## 5. Diagnóstico

### 5.1. Top 10 Diagnóstico

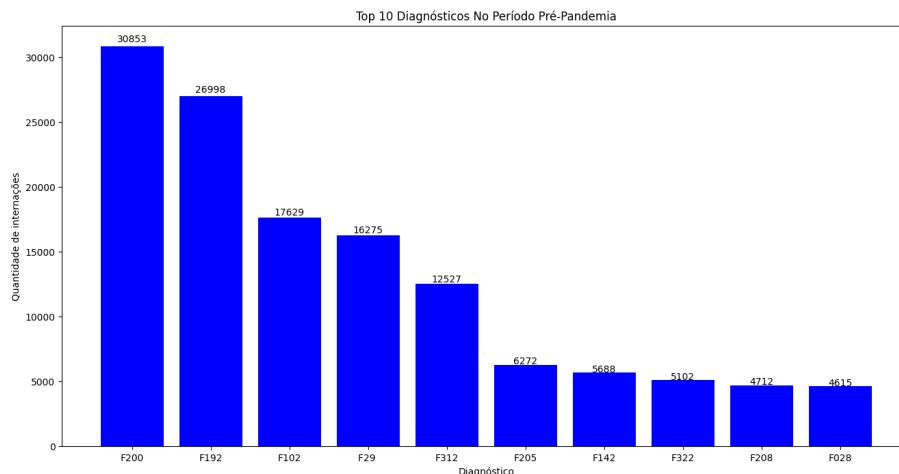
- **1º F200:** Esquizofrenia paranóide
- **2º F192:** Transtornos mentais e comportamentais devidos ao uso de múltiplas drogas e ao uso de outras substâncias psicoativas - síndrome de dependência
- **3º F102:** Transtornos mentais e comportamentais devidos ao uso de álcool - síndrome de dependência
- **4º F29:** Psicose não-orgânica não especificada
- **5º F312:** Transtorno afetivo bipolar, episódio atual maníaco com sintomas psicóticos
- **6º F205:** Esquizofrenia residual
- **7º F142:** Transtornos mentais e comportamentais devidos ao uso da cocaína - síndrome de dependência
- **8º F322:** Episódio depressivo grave sem sintomas psicóticos
- **9º F208:** Outras esquizofrenias
- **10º F028:** Demência em outras doenças especificadas classificadas em outra parte

O ranking é liderado por esquizofrenia (F200) e dependência de múltiplas drogas (F192), evidenciando que o sistema foca em casos de altíssima complexidade. A forte presença de álcool e cocaína (F102, F142) indica uma crise de dependência química que sobrecarrega os leitos hospitalares. Psicoses e episódios maníacos (F29, F312) reforçam que a internação é usada como último recurso para conter crises agudas. Esse perfil de diagnósticos graves revela a



# Documentação Análise Exploratória de Dados

carência de suporte preventivo na rede básica de saúde. Assim, o topo da lista expõe um cenário de contenção de danos em patologias severas e crônicas.



## 6. Análise Geográfica

### 6.1. Fluxo de Origem-Destino

39% dos pacientes são obrigados a viajar para se tratar, provando que a saúde mental em SP é centralizada e desigual. Isso isola o paciente da família e sobrecarrega as cidades grandes, revelando que a rede local de muitas cidades é insuficiente ou inexistente.

### 6.2. Municípios de Residência vs Internação

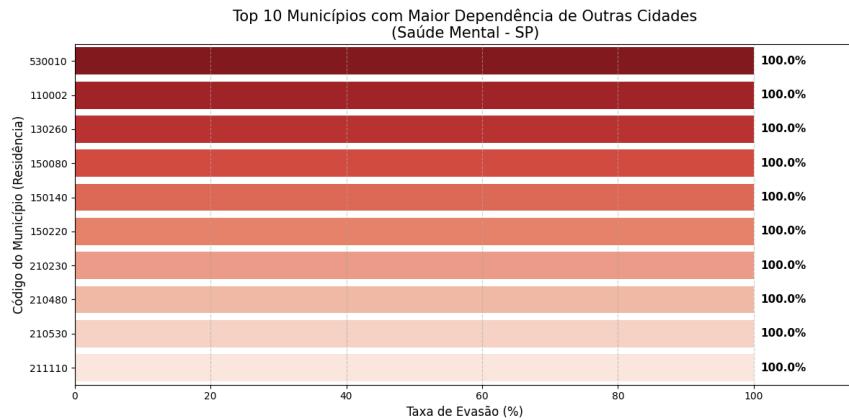
Observa-se que diversos municípios apresentam 100% de evasão nas internações em saúde mental, indicando ausência de atendimento local. Em alguns casos, o volume absoluto é baixo; contudo, há municípios com mais de 30 ou 40 internações totalmente realizadas fora do território de residência, o que evidencia dependência estrutural de cidades-polo e possível desigualdade na distribuição regional dos serviços especializados.

#### 6.2.1. Cidades fora de SP que mais dependem de outros municípios

Existem municípios com 100% de taxa de evasão, o que significa que nenhum paciente consegue ser atendido em sua própria cidade. Esse dado revela um vazio assistencial total nessas localidades, onde a dependência de outros municípios é absoluta devido à falta completa de leitos ou hospitais psiquiátricos para a população local.



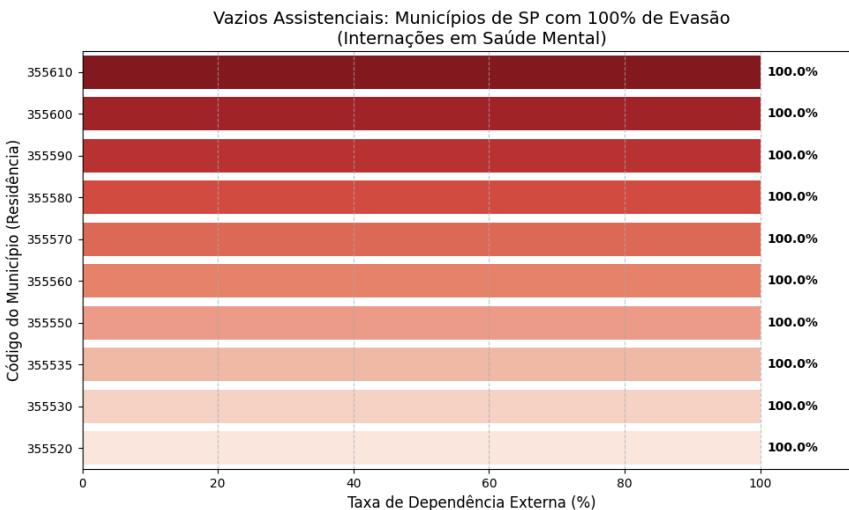
## Documentação Análise Exploratória de Dados



### 6.2.2. Cidades de SP que mais dependem de outros municípios

Diversos municípios paulistas apresentam 100% de taxa de evasão, o que caracteriza um vazio assistencial absoluto na rede de saúde mental.

Nessas localidades, não existe nenhuma estrutura própria para internação, o que torna a dependência de cidades vizinhas total e obriga 100% dos pacientes a buscarem socorro hospitalar fora de seu local de residência.

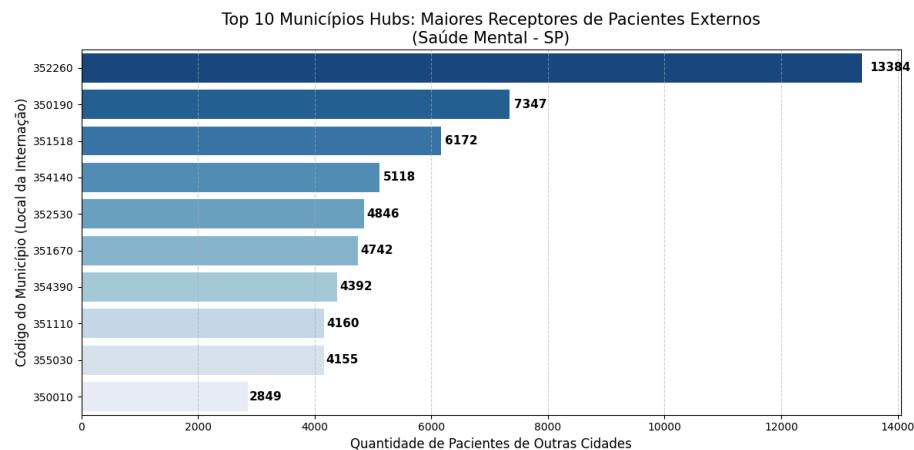


### 6.2.3. Cidades que mais recebem pacientes de outros municípios

Enquanto muitos municípios sofrem com a falta de leitos, cidades como **Itu (352260)** e **Amparo (350190)** atuam como grandes polos receptores, absorvendo a demanda de toda a sua região. O destaque fica para Itu, que recebeu sozinha mais de 13 mil pacientes de outras localidades, confirmando que a rede de saúde mental paulista é sustentada por alguns centros de referência que sobrecarregam sua estrutura para suprir os vazios assistenciais das cidades vizinhas.



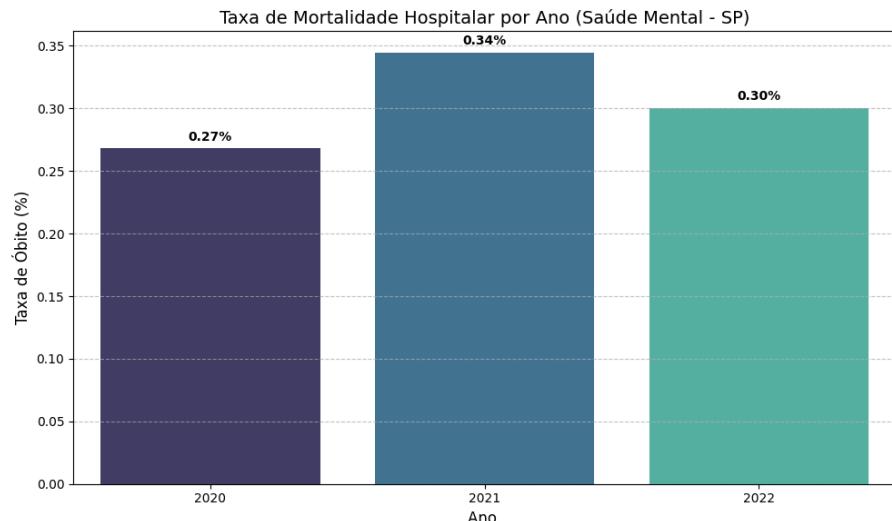
# Documentação Análise Exploratória de Dados



## 7. Óbito

### 7.1. Taxa de óbito por ano

A análise da taxa de mortalidade durante a pandemia revela que o ano de 2021 foi o período mais crítico, atingindo o pico de 0,34%. Esse aumento reflete o agravamento dos desfechos clínicos no momento mais severo da crise sanitária, seguido por uma redução para 0,30% em 2022, patamar que ainda permanece superior ao registrado no início da série em 2020 (0,27%).



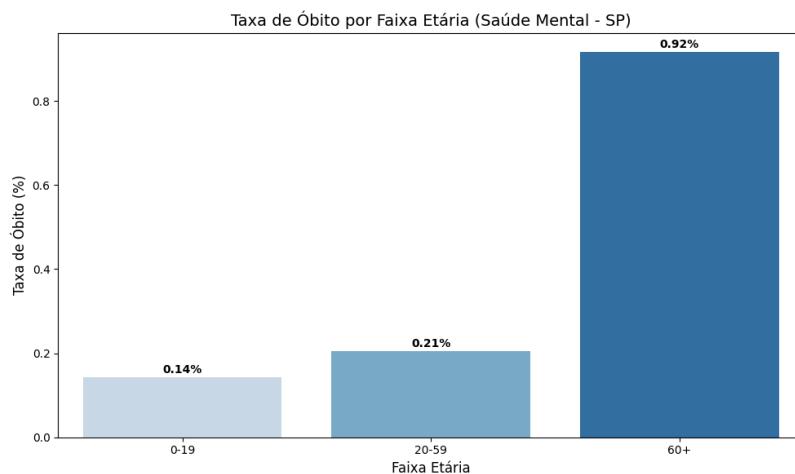
### 7.2. Óbito por idade

Esta análise das taxas de óbito por faixa etária durante a pandemia reforça que a idade é o principal fator de risco para a letalidade hospitalar em saúde mental. Enquanto as faixas de 0-19 anos (0,14%) e 20-59 anos (0,21%) mantiveram índices relativamente baixos, a população com 60 anos ou mais apresentou uma taxa de 0,92%, um valor significativamente superior aos demais grupos. Essa disparidade



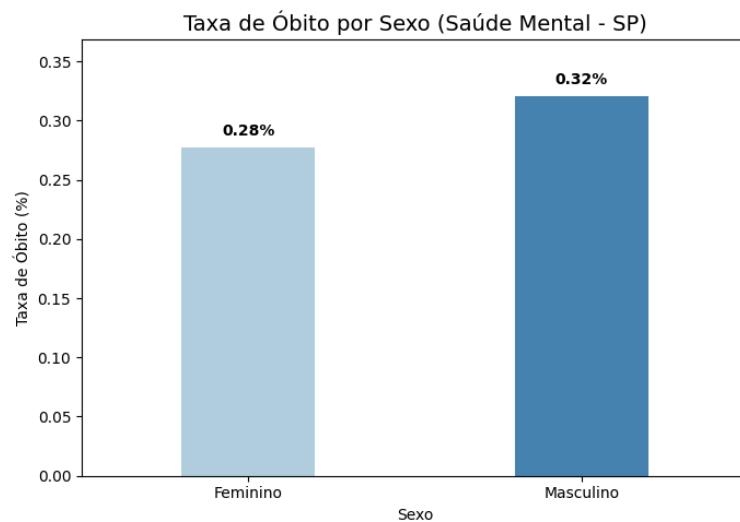
## Documentação Análise Exploratória de Dados

evidencia a alta vulnerabilidade dos idosos no sistema hospitalar durante o período crítico da crise sanitária.



### 7.3. Óbito por sexo

A análise da letalidade hospitalar por gênero durante a pandemia revela que o sexo Masculino apresentou um índice de 0,32%, sendo proporcionalmente superior ao registrado pelo sexo Feminino (0,28%). Essa diferença reforça que, além do fator idade, o gênero masculino constituiu um grupo de maior vulnerabilidade para desfechos fatais em internações de saúde mental no estado de São Paulo durante o período de crise sanitária.



### 7.4. Óbito por raça/cor

A análise por Raça/Cor mostra que o grupo Amarela registrou a maior letalidade hospitalar no período pandêmico, com uma taxa de 0,74%, seguido pelo grupo de registros Ignorados (0,41%). Esses valores são consideravelmente



## Documentação Análise Exploratória de Dados

superiores aos observados nos grupos Preta (0,31%), Branca (0,29%) e Parda (0,24%). A elevada taxa no grupo "Ignorado" acende um alerta para a qualidade do preenchimento das fichas hospitalares, enquanto o índice do grupo Amarela sugere a necessidade de uma investigação mais profunda sobre vulnerabilidades específicas dessa população durante a crise sanitária.

