



EDA Pré-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 **711.022 entradas**, cada uma composta por **23 variáveis** distintas para análise.

1.2. Colunas do Dataset

O conjunto de dados consolidado para o período pré-pandemia contém **711.022 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

Consistência da Base: O dataset final não apresenta valores nulos (missing values) em nenhuma das 23 colunas analisadas.

1.5. Valores Duplicados



Documentação Análise Exploratória de Dados

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 **711.022 registros**.

Identificação de duplicatas: foram identificados **18.913 registros duplicados** (aproximadamente **2,65%** 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, 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 692.109 internações únicas, assegurando que cada evento assistencial fosse contabilizado apenas uma vez.

1.6. Recorte Temporal

O período analisado foi de janeiro de 2014 a dezembro de 2019.

2. Panorama Geral

2.1. Número total de internações

Após o processo de limpeza e desduplicação, o estudo consolidou um universo de **692.109 internações únicas** no estado de São Paulo entre 2014 e 2019.

2.2. Média Mensal de Internações

O número médio de internações por mês no período analisado foi de aproximadamente 9.612 internações, calculado a partir da contagem mensal de internações únicas.

2.3. Quantidade de municípios residente e de internação

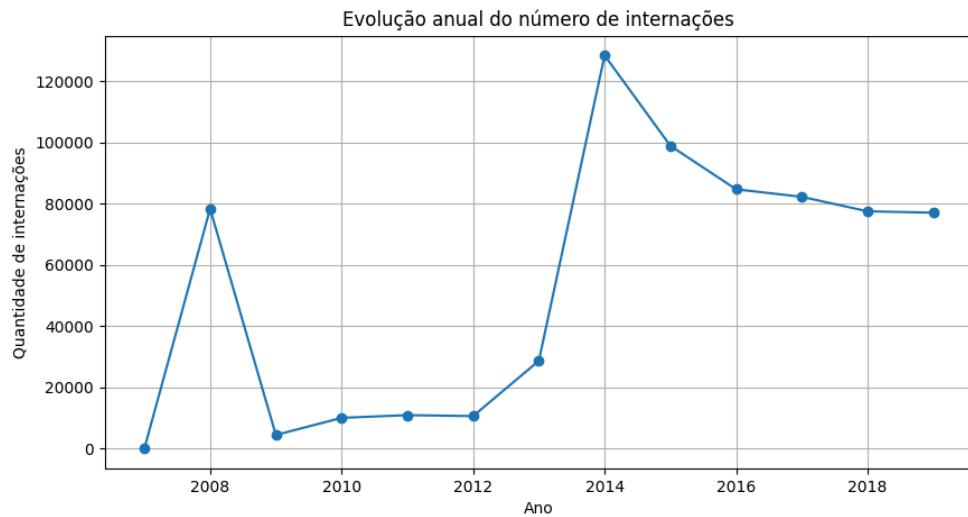
Enquanto os pacientes são provenientes de **1.038 municípios** diferentes, a rede hospitalar está concentrada em apenas **315 cidades**. Essa disparidade revela uma centralização dos serviços de saúde mental, obrigando a maioria dos pacientes a se deslocarem de suas cidades de origem para obter assistência hospitalar, evidenciando uma desigualdade no acesso à saúde.

3. Análise Temporal



3.1. Evolução Anual das Internações

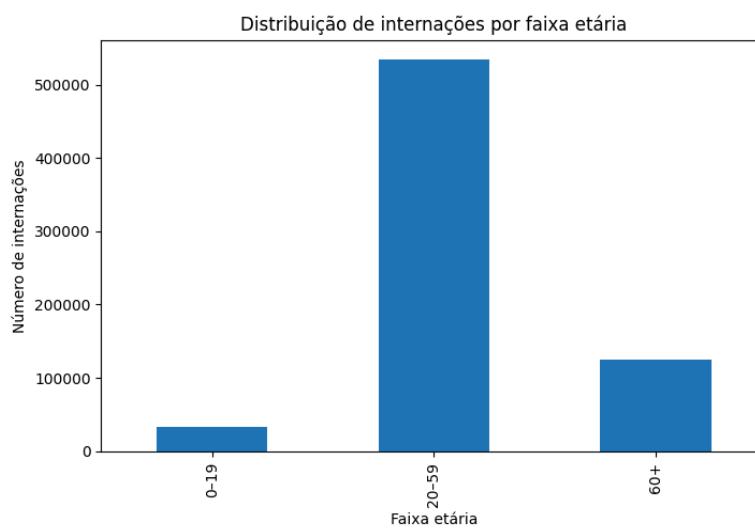
As internações psiquiátricas em São Paulo cresceram rápido até atingirem o seu ponto máximo em **2014**. Depois disso, o número de pacientes começou a cair de forma constante todos os anos até 2019. Essa queda sugere que houve uma mudança na organização dos hospitais ou uma redução das vagas disponíveis antes mesmo da pandemia começar, 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 quatro vezes o total de idosos (60+). Destaca-se a vida adulta como o período de maior vulnerabilidade ou registro de crises agudas.

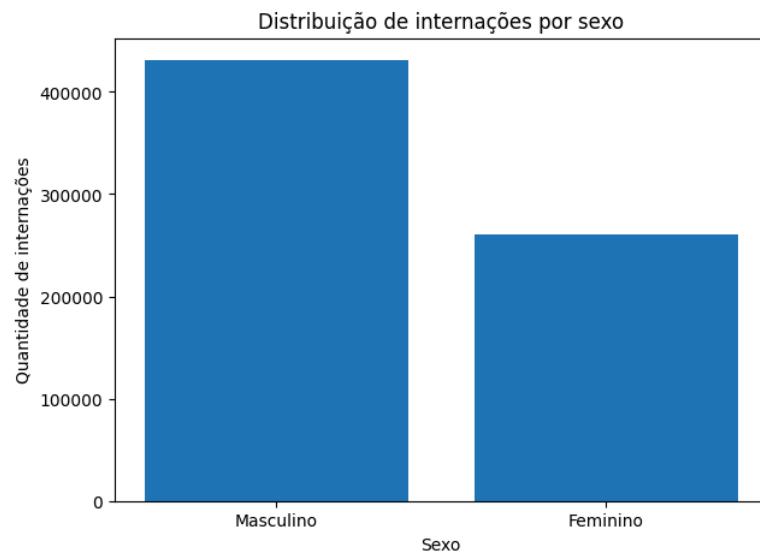


4.2. Distribuição de Sexo



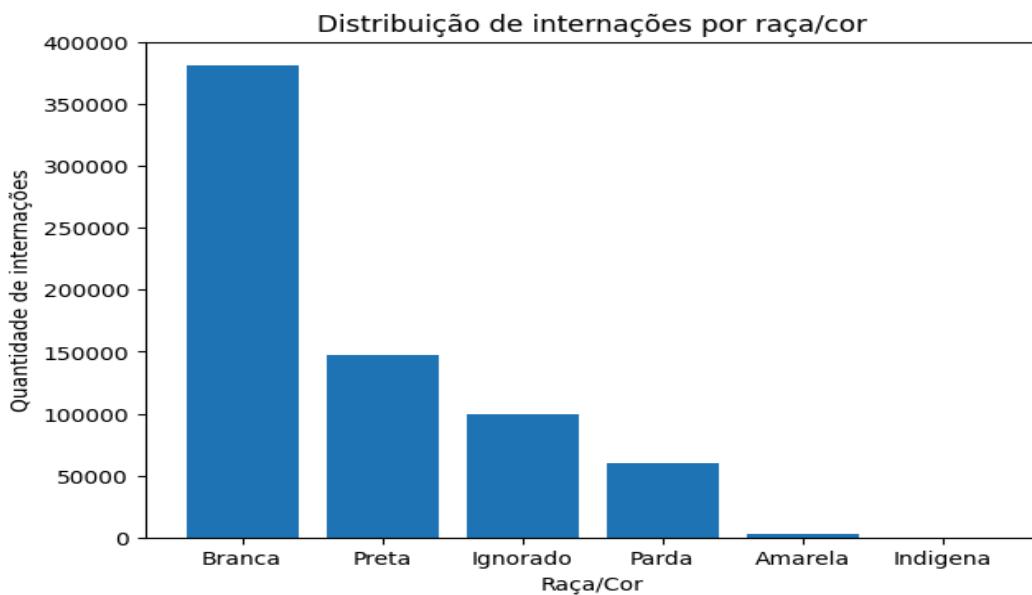
Documentação Análise Exploratória de Dados

O público **masculino** representa a maioria expressiva das internações (**62%**), com um volume cerca de 1,6 vez maior que o feminino, conforme o gráfico abaixo. 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** (**55%**) contrasta com a soma de **pretos e pardos** (**30%**), sugerindo disparidades no acesso ou no perfil das internações. Chama a atenção o alto volume de dados ignorados (**14,4%**), o que prejudica a precisão das políticas de equidade, enquanto a baixa presença de indígenas e amarelos aponta para uma possível exclusão ou subnotificação severa desses grupos no sistema.



5. Diagnóstico

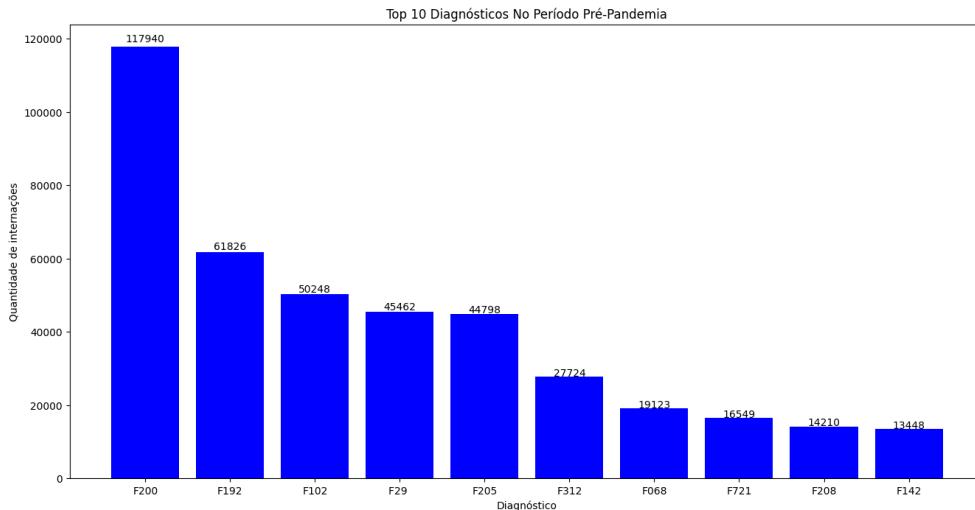
5.1. Top 10 Diagnóstico

A análise do período pré-pandemia revela que o perfil das internações em São Paulo era dominado por quadros de **psicoses e esquizofrenia**, além de uma forte presença de **transtornos decorrentes do uso de substâncias** (álcool, cocaína e múltiplas drogas).

O destaque do **F29 (Psicoses não orgânicas)** no topo do ranking aponta para uma alta frequência de diagnósticos genéricos ou crises agudas que exigiam hospitalização imediata. A presença significativa de **dependência química** (3º ao 5º lugar) reforça que, antes da COVID-19, a rede hospitalar era o principal suporte para casos graves de abuso de substâncias, enquanto o grupo das **esquizofrenias** (6º ao 8º lugar) evidencia o caráter crônico da demanda assistencial da época.



Documentação Análise Exploratória de Dados



6. Análise Geográfica

6.1. Fluxo de Origem-Destino

Quase 40% 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

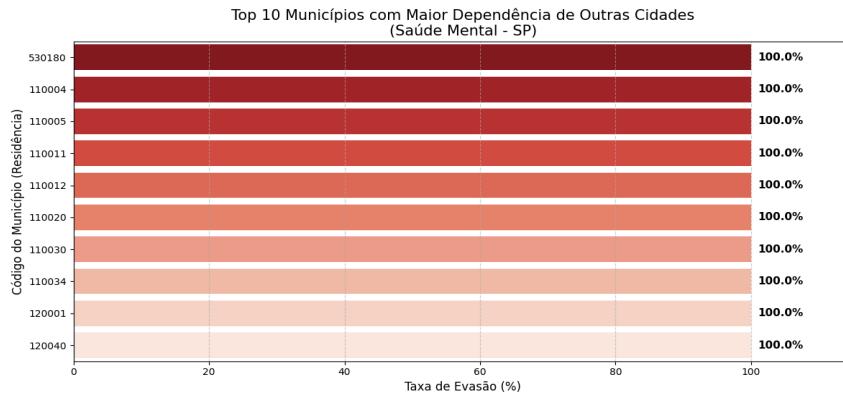
O fluxo de internações revela uma dependência crítica de "cidades-hospital", como Itu e Jaú, que absorvem a demanda de todo o estado. O fato de a própria Capital ser a maior "exportadora" de pacientes prova que a rede local é insuficiente, forçando deslocamentos longos que isolam o paciente de sua família e evidenciam a centralização da assistência em São Paulo.

6.2.1. Cidades que mais dependem de outros municípios

A presença de pacientes de outros estados (RO, AC, DF) com 100% de evasão destaca São Paulo como um polo de referência nacional, atraindo casos graves que não encontram suporte em suas regiões de origem.

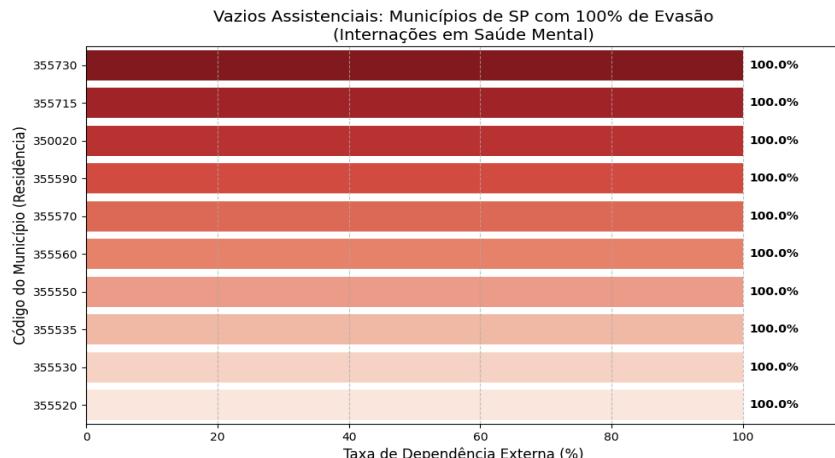


Documentação Análise Exploratória de Dados



6.2.2. Cidades de SP que mais dependem de outros municípios

A análise das cidades como Estrela d'Oeste, Zacarias e Adolfo revela vazios assistenciais absolutos: com 100% de evasão, nenhum morador é atendido onde vive. Isso prova que, no interior de SP, a saúde mental é desigual e excludente, obrigando o paciente de cidades pequenas a se isolar da família e buscar socorro em polos distantes por falta de qualquer estrutura local.

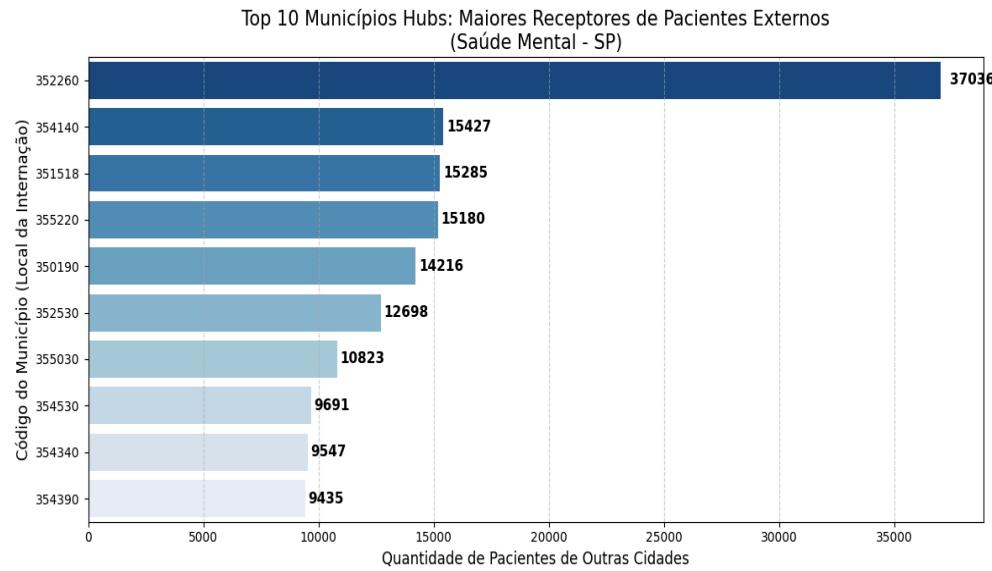


6.2.3. Cidades que mais recebem pacientes de outros municípios

Municípios como Itu (líder absoluto com 37 mil pacientes externos), Presidente Prudente e Sorocaba atuam como os grandes "pulmões" do sistema, absorvendo a demanda de todo o estado. Essa hiper-centralização prova que a assistência hospitalar não é regionalizada: enquanto a maioria das cidades não oferece nada, esses poucos polos carregam a sobrecarga, forçando o deslocamento em massa de pacientes e evidenciando a profunda desigualdade na distribuição de leitos.



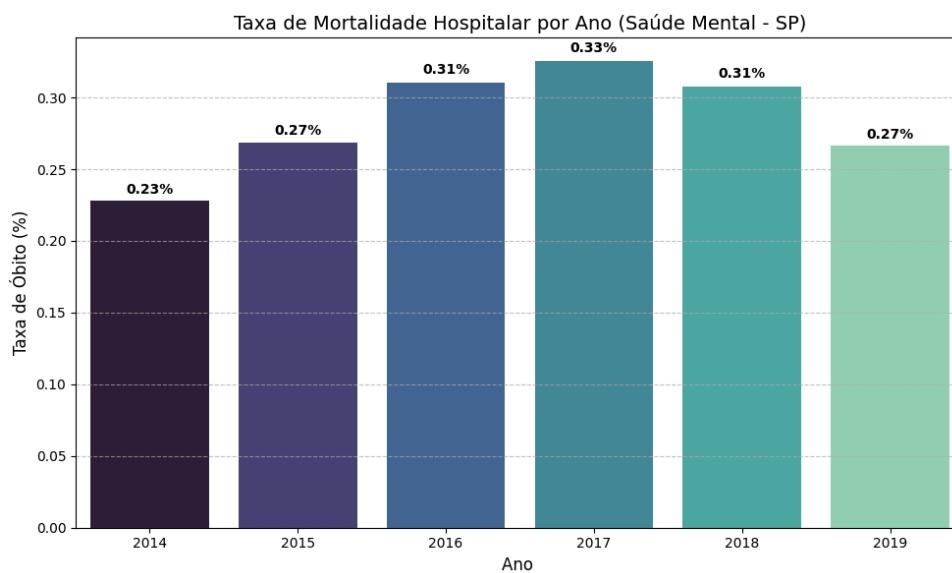
Documentação Análise Exploratória de Dados



7. Óbito

7.1. Taxa de óbito por ano

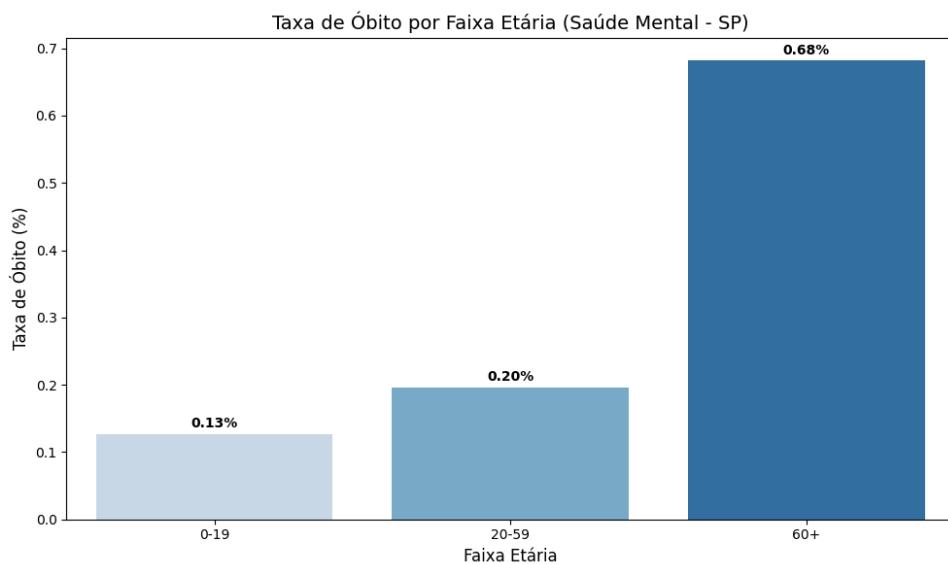
Os dados revelam que a mortalidade hospitalar em saúde mental no estado de São Paulo já apresentava uma trajetória de ascensão anos antes da crise sanitária de 2020. O pico observado em 2017 (0,33%) acende um alerta sobre a precarização do cuidado clínico dentro das unidades psiquiátricas, sugerindo que o sistema já operava sob pressão ou com falhas no suporte à saúde física dos pacientes internados.



7.2. Óbito por idade

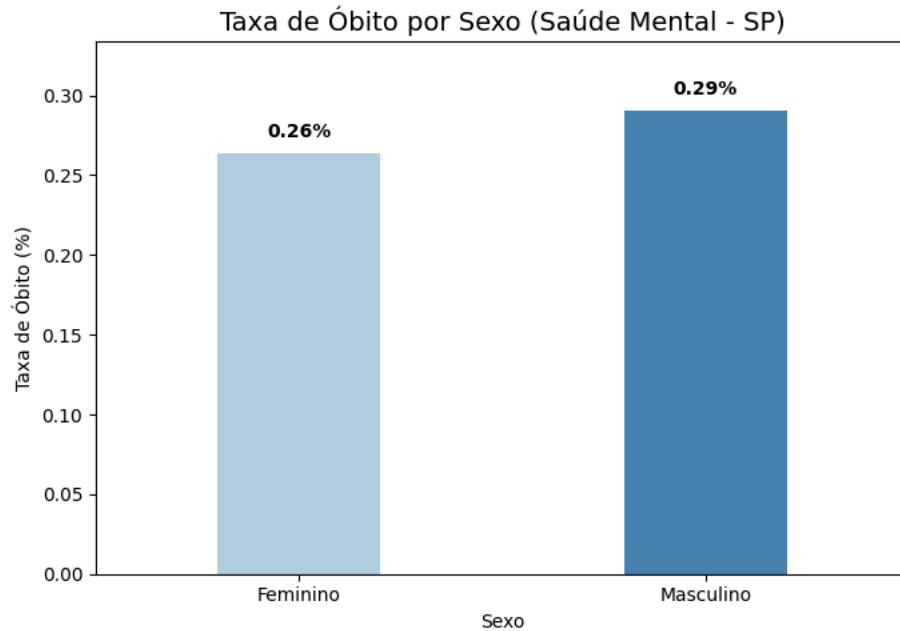


A análise por faixa etária revela que a mortalidade hospitalar é fortemente influenciada pela idade, com um risco mais de três vezes superior para os idosos (60+), que apresentam uma taxa de 68%. Enquanto os grupos de 0-19 anos (0,13%) e 20-59 anos (0,20%) mantêm índices mais baixos, o salto na terceira idade evidencia a vulnerabilidade clínica extrema desse público e a complexidade de tratar comorbidades físicas associadas a transtornos mentais no ambiente hospitalar.



7.3. Óbito por sexo

A análise do gráfico de sexo revela que a taxa de óbito hospitalar em saúde mental é ligeiramente superior entre o público masculino (0,29%) em comparação ao feminino (0,26%). Embora a diferença seja pequena, ela sugere que as internações de homens podem estar associadas a quadros clínicos de maior gravidade ou comorbidades físicas mais severas, frequentemente ligadas ao uso abusivo de substâncias.



7.4. Óbito por raça/cor

A análise revela disparidades importantes, com a população Amarela (0,42%) apresentando a maior taxa. Um ponto crítico é o grupo "Ignorado" (0,38%), que possui a segunda maior letalidade; isso indica que a falha no preenchimento de dados (subnotificação) pode estar ocultando a realidade de pacientes em situações de maior vulnerabilidade social.

