



# **ESTUDO DE CASO: SUICÍDIOS NA ÉPOCA DA PANDEMIA COVID-19**

**DAVID BATISTA  
GABRIEL ALVES  
MARIA EDUARDA BOVO**

# PROPOSTA DO PROJETO

Analisar a base de dados do DATASUS aplicando algoritmos de agrupamentos para geração de conhecimento não explícito sobre suicídios em 2022. Investigar padrões de mortalidade por lesões autoprovocadas intencionalmente (CID-10: X60-X84), problema de saúde pública agravado pelo aumento de 25% em ansiedade e depressão pós-pandemia (OPAS/OMS, 2022). A aplicação de clustering identificará perfis de risco e padrões ocultos, subsidiando políticas públicas de prevenção baseadas em evidências.



# FONTE (DATASUS)



- O Ministério da Saúde (MS) disponibiliza o DATASUS, o principal portal de dados do Sistema Único de Saúde (SUS) para consulta pública.
- Os dados de mortalidade por suicídio são definidos pela Classificação Internacional de Doenças (CID-10), especificamente os códigos X60 a X84 (Lesões Autoprovocadas Intencionalmente).
- O total de atributos do dataset é de 17 atributos, que contem raça, causa, estado, estado civil, ano e etc, contendo dados faltantes, categóricos e néricos

# TRATAMENTO DOS DADOS

- O total de atributos do dataset bruto é de 17 atributos, que contém raça, causa(CID), estado, estado civil, ano e etc, contendo dados faltantes, categóricos e numéricos.
- Remoção de colunas: 'ESC2010', 'ESCFALAGR1', 'CODMUNRES', 'LOCOCOR', 'NATURAL', 'CODMUNOCOR', 'OCUP', por não apresentar ganho de conhecimento a análise.
- Foi removido todos os registros com valores nulos.
- Dataset filtrado para os registros depois de 2020.

# BASE DE DADOS BRUTA

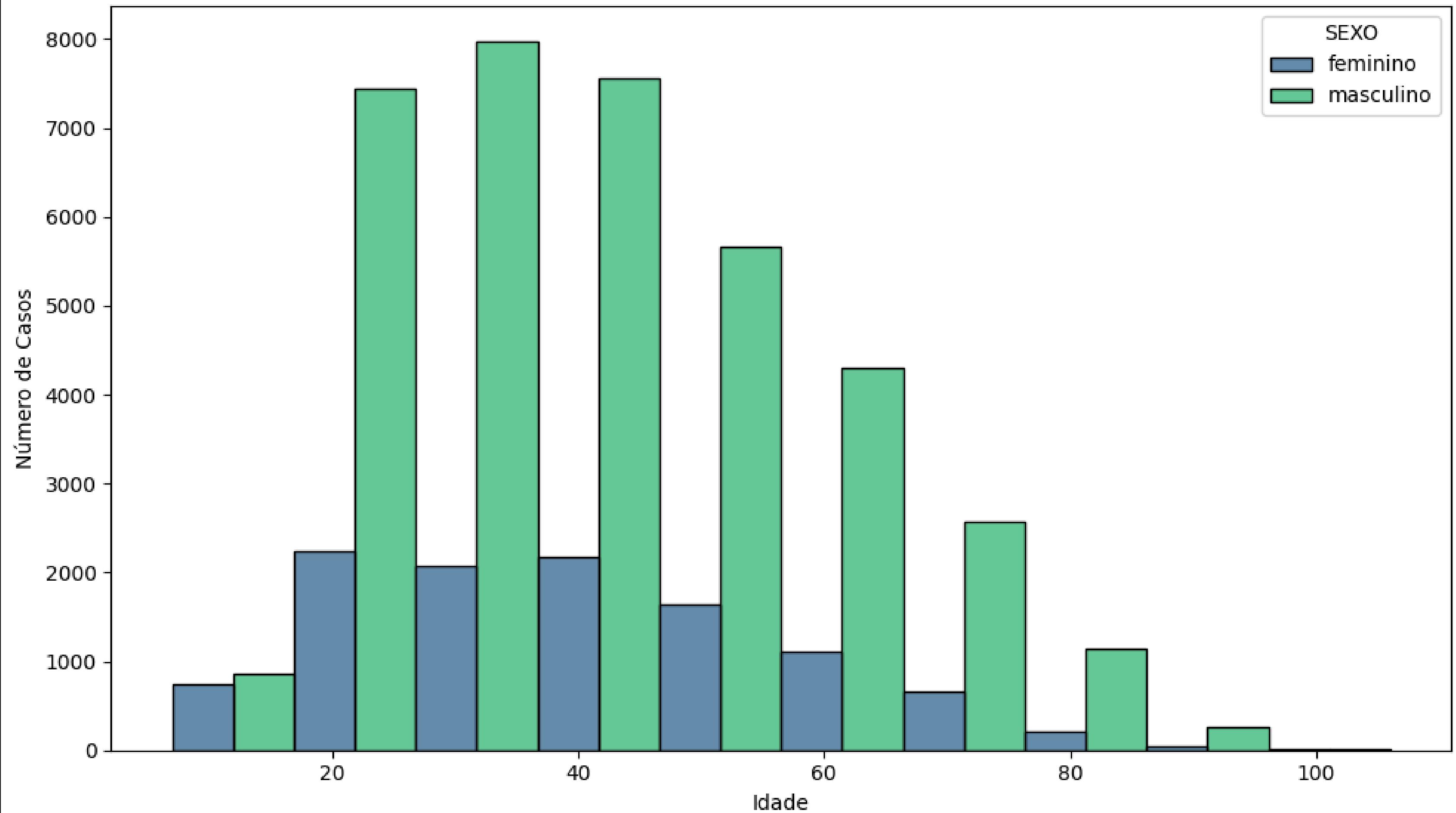
	ESTADO	ANO	DTOBITO	NATURAL	DTNASC	IDADE	SEXO	RACACOR	ESTCIV	ESC	OCUP	CODMUNRES	LOCOCOR	CODMUNOCOR	CAUSABAS	ESC2010
0	PR	1996	1996-10-24	841.0	1963-08-05	33.0	masculino	NaN	Solteiro	Nenhuma	61200.0	Pinhal de São Bento	outros	Pinhal de São Bento	X709	NaN
1	PR	1996	1996-02-04	152.0	1920-09-21	75.0	masculino	NaN	Casado	Nenhuma	62100.0	Assaí	domicílio	Assaí	X680	NaN
2	PR	1996	1996-01-17	841.0	1961-08-28	34.0	masculino	NaN	Solteiro	NaN	62100.0	Abatiá	hospital	Abatiá	X680	NaN
3	PR	1996	1996-02-07	841.0	1977-09-24	18.0	masculino	NaN	Solteiro	Nenhuma	62100.0	Santa Cecília do Pavão	hospital	São Sebastião da Amoreira	X680	NaN
4	PR	1996	1996-02-09	841.0	1956-08-30	39.0	masculino	NaN	União consensual	NaN	62100.0	Santa Mariana	domicílio	Santa Mariana	X700	NaN
5	PR	1996	1996-02-12	841.0	1970-12-04	25.0	masculino	NaN	Solteiro	NaN	62100.0	Santa Mariana	domicílio	Santa Mariana	X700	NaN
6	PR	1996	1996-01-04	841.0	1973-04-15	22.0	masculino	NaN	Solteiro	8 a 11 anos	39300.0	Cornélio Procópio	hospital	Cornélio Procópio	X628	NaN

# BASE DE DADOS LIMPA

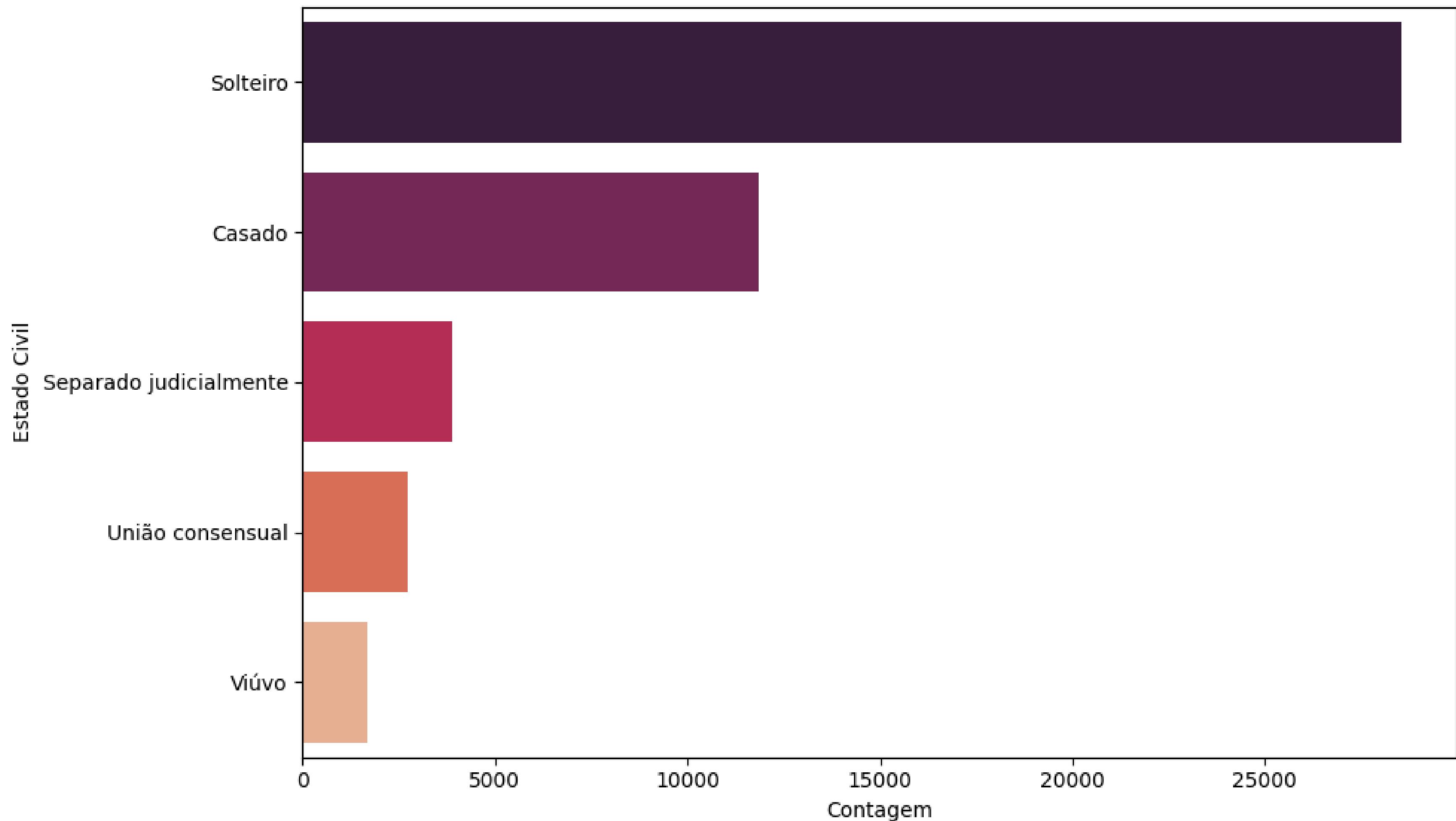
ESTADO	ANO	DTOBITO	DTNASC	IDADE	SEXO	RACACOR	ESTCIV	ESC	CAUSABAS	MES_OBITO
PR	2020	2020-06-12	1958-07-12	61.0	feminino	Branca	Casado	12 anos e mais	X700	6
PR	2020	2020-06-13	1996-08-08	23.0	masculino	Branca	Solteiro	12 anos e mais	X700	6
PR	2020	2020-04-27	1942-05-06	77.0	masculino	Branca	Casado	4 a 7 anos	X720	4
PR	2020	2020-05-18	1999-01-31	21.0	feminino	Branca	Solteiro	8 a 11 anos	X709	5
PR	2020	2020-04-15	1976-03-11	44.0	masculino	Parda	Separado judicialmente	4 a 7 anos	X700	4
PR	2020	2020-04-16	1974-06-11	45.0	masculino	Parda	Solteiro	4 a 7 anos	X700	4

# ALGUNS INSIGHTS

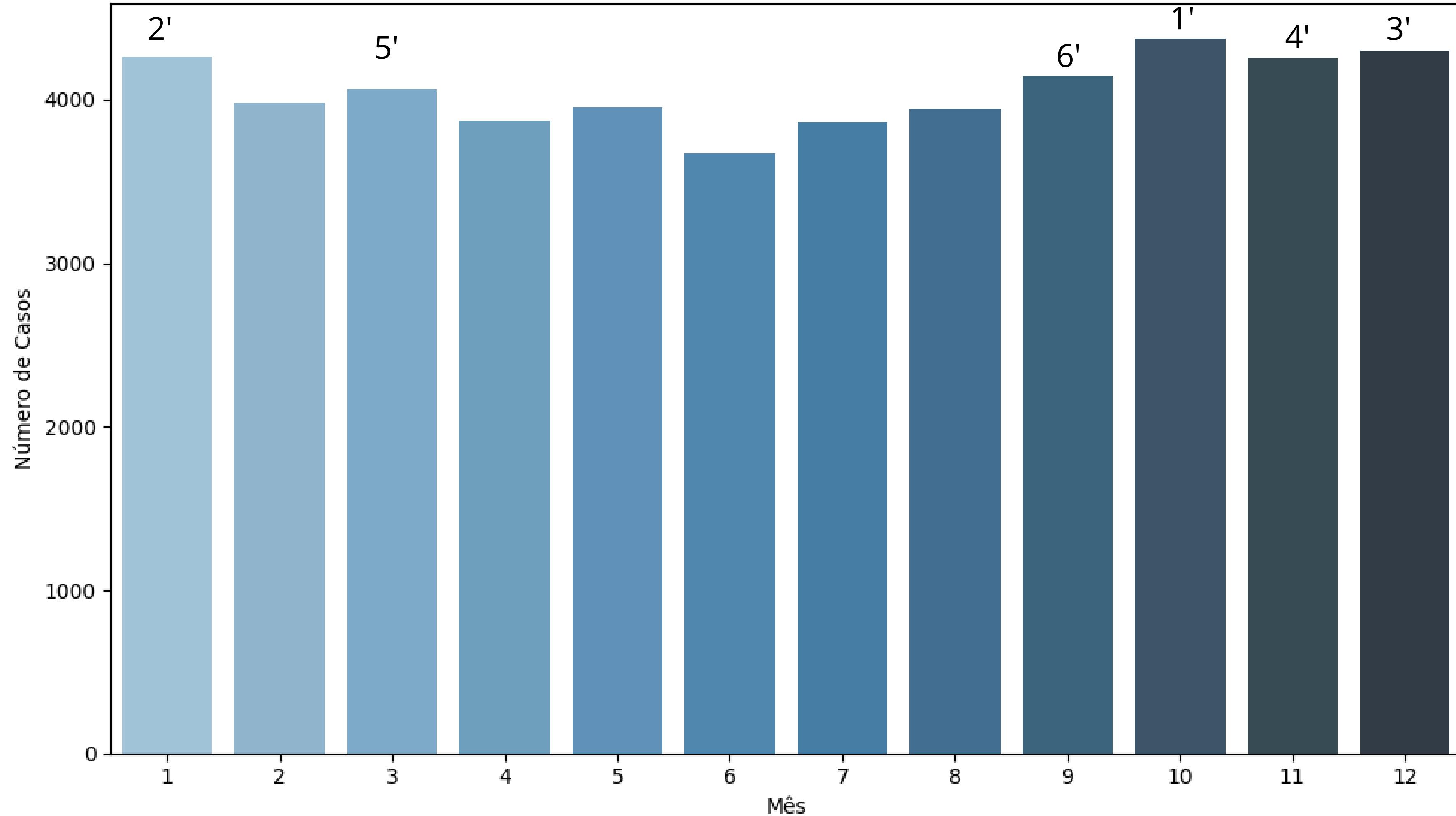
## Distribuição de Casos por Faixa Etária e Sexo - 2020 em diante



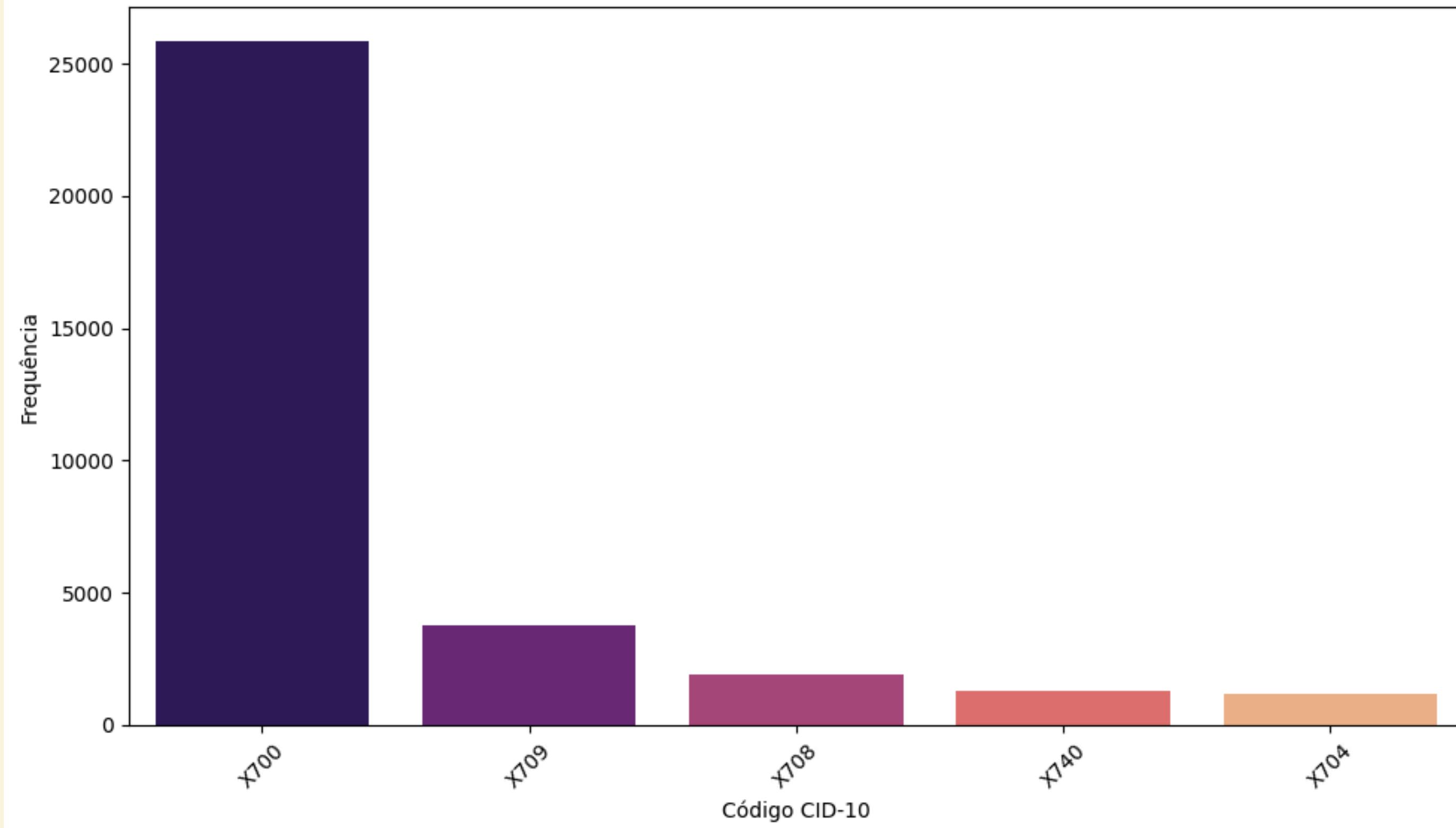
## Ocorrências por Estado Civil - 2020 em diante



## Sazonalidade: Ocorrências por Mês - 2020 em diante



### Top 5 Códigos CID (Causa Básica) - 2020 em diante



**X700:** Asfixia (Enforcamento/Estrangulamento) - Ocorrido em residência.

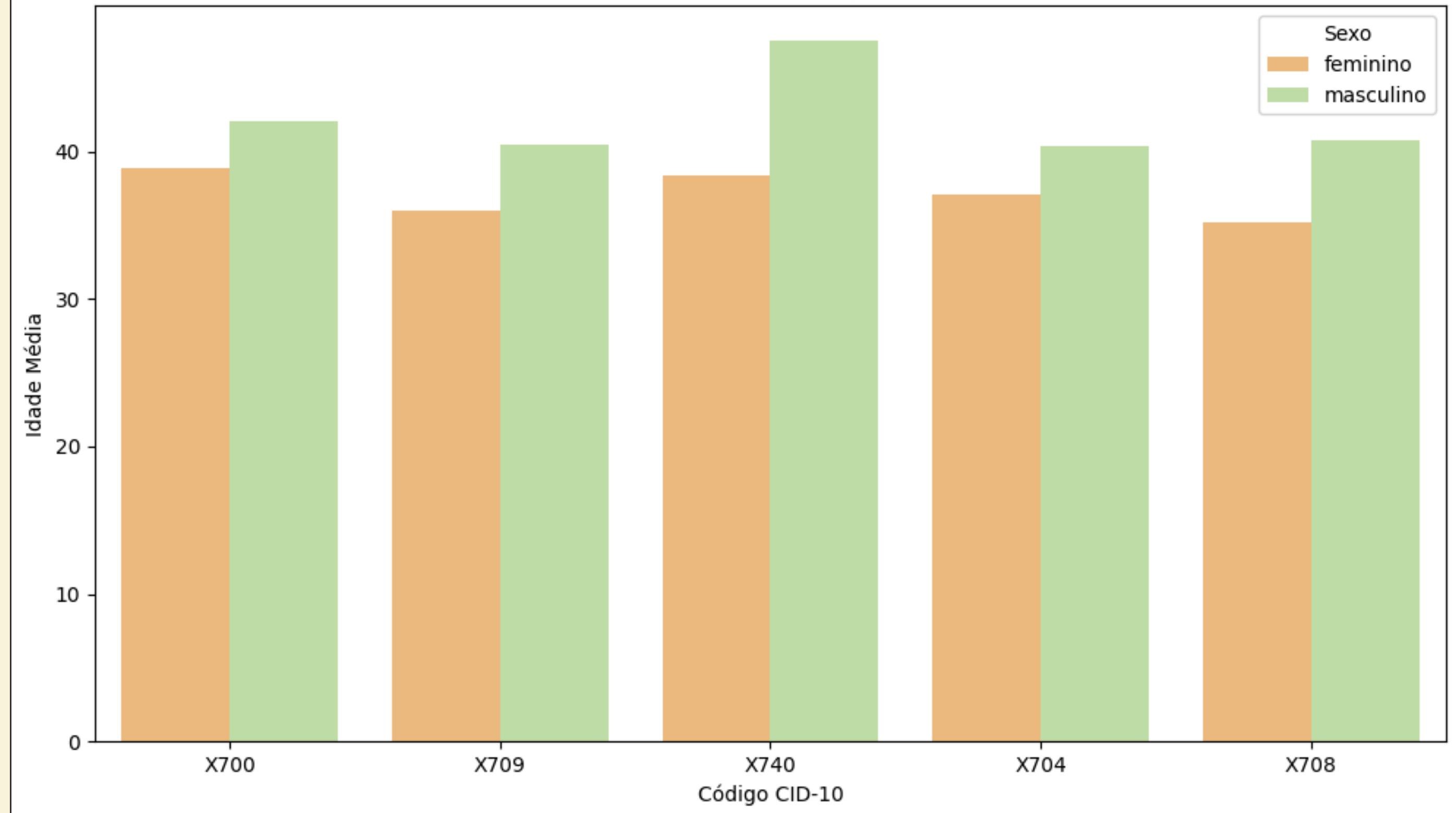
**X709:** Asfixia (Enforcamento/Estrangulamento) - Ocorrido em local não especificado.

**X740:** Disparo de Arma de Fogo - Ocorrido em residência

**X708:** Disparo de Arma de Fogo - Ocorrido em local não especificado.

**X749:** Asfixia (Enforcamento/Estrangulamento) - Ocorrido em outros locais especificados.

### Média de Idade por Método e Gênero (Top 5 Causas) - 2020 em diante



**X700:** Asfixia (Enforcamento/Estrangulamento) - Ocorrido em residência.

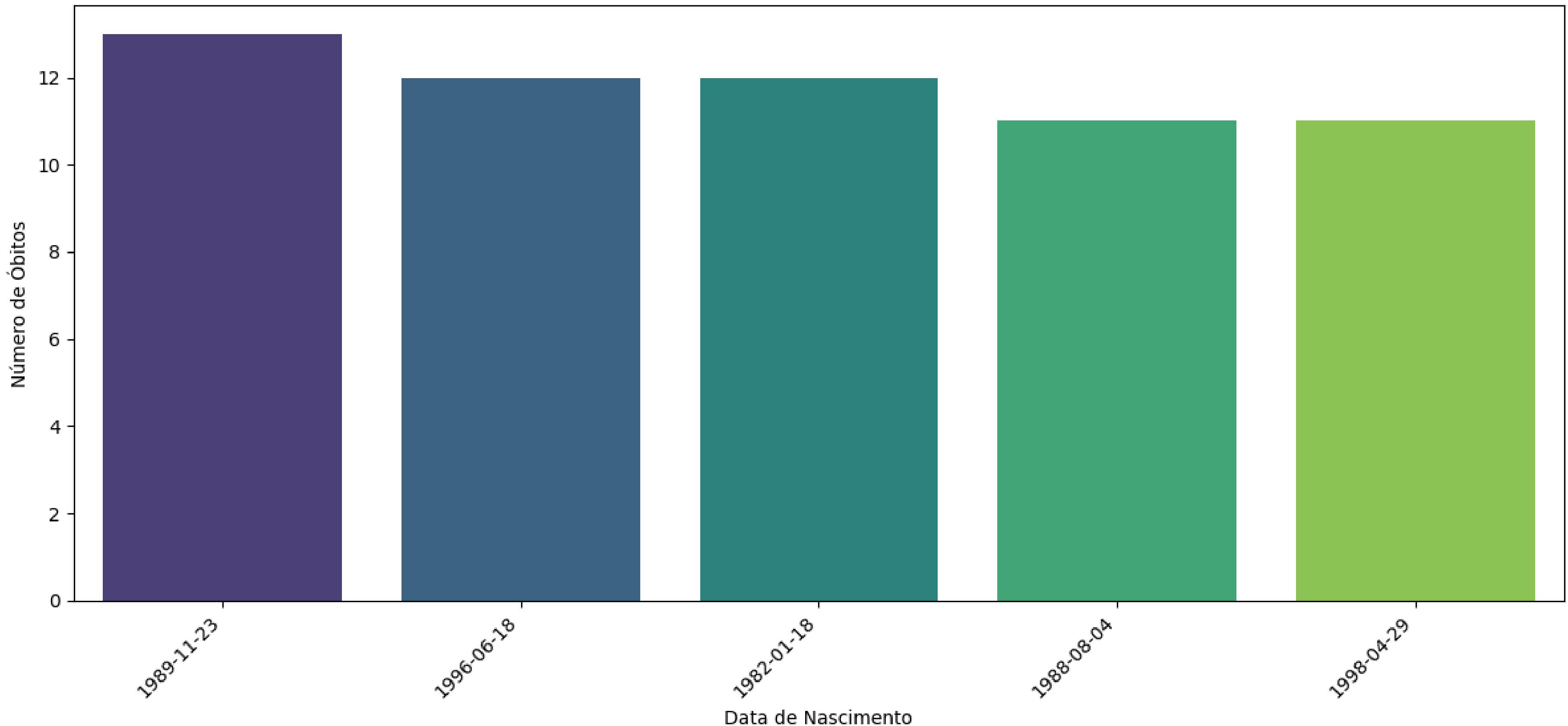
**X709:** Asfixia (Enforcamento/Estrangulamento) - Ocorrido em local não especificado.

**X740:** Disparo de Arma de Fogo - Ocorrido em residência

**X708:** Disparo de Arma de Fogo - Ocorrido em local não especificado.

**X749:** Asfixia (Enforcamento/Estrangulamento) - Ocorrido em outros locais especificados.

## Top 5 Datas de Nascimento com Mais Óbitos - 2020 em diante



# **AGRUPAMENTO**

# ATRIBUTOS SELECIONADOS PARA ANÁLISE

**Para definir os grupos, o algoritmo analisou as seguintes características de cada registro:**

- Variáveis Numéricas:
  - Idade: Faixa etária do indivíduo.
  - Ano e Mês do Óbito: Para identificar sazonalidade temporal.
- Variáveis Categóricas (Qualitativas):
  - Sexo: Masculino ou Feminino.
  - Raça/Cor: Branca, Parda, Preta, etc.
  - Estado Civil: Solteiro, Casado, Viúvo, etc.
  - Escolaridade (ESC): Nível de instrução.

# PREPARAÇÃO DA BASE DE DADOS

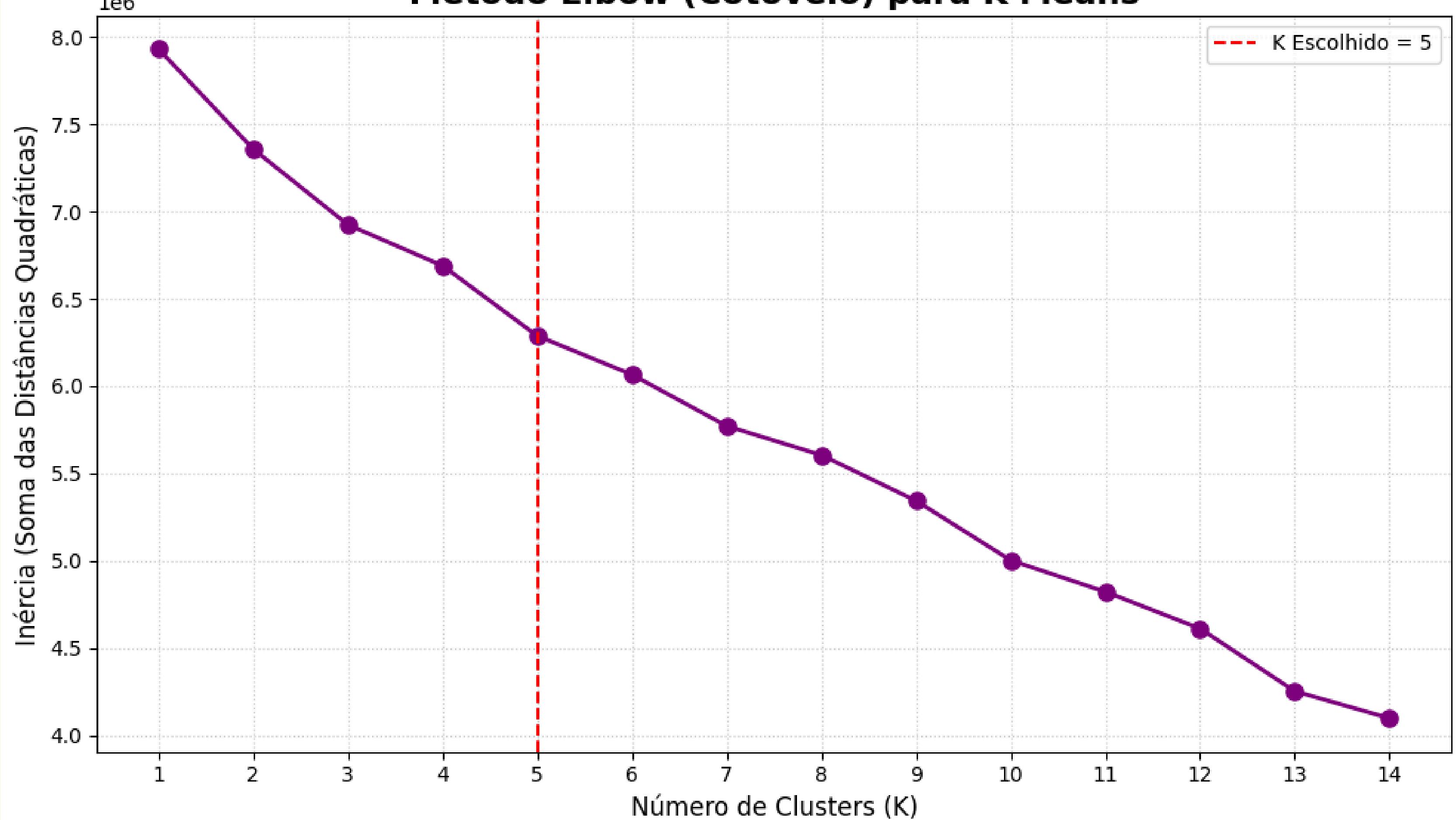
**Antes de rodar o algoritmo, os dados precisaram ser transformados em tipos numéricos:**

1. Limpeza: Preenchimento de alguns dados faltantes (usando a média para idades e a moda para categorias).
2. One-Hot Encoding: Transformação de textos em números.
  - Exemplo: A coluna "Sexo" vira duas colunas numéricas: "Sexo\_M" (0 ou 1) e "Sexo\_F" (0 ou 1).
3. Padronização (StandardScaler): Colocamos todos os dados na mesma escala.
  - Sem isso, a "Idade" (que vai de 0 a 100) teria um peso muito maior na decisão do algoritmo do que o "Mês" (que vai de 1 a 12). A padronização equilibra o peso de todas as variáveis.

# O MÉTODO ELBOW (COTOVELO)

- Métrica (Inércia): Mede o quanto "compactos" são os grupos. Quanto menor o valor, mais coesos são os grupos.
- Lógica do Cotovelo:
  - No início, aumentar o número de grupos ( $K$ ) diminui drasticamente o erro (inércia).
  - Chega um ponto onde criar mais grupos não traz um ganho significativo de qualidade. Esse ponto de "retorno decrescente" é o cotovelo.
- Decisão: Escolhemos  $K = 5$ . É o ponto de equilíbrio onde temos perfis distintos o suficiente para análise, sem fragmentar excessivamente os dados.

## Método Elbow (Cotovelo) para K-Means



# PERFIS IDENTIFICADOS

## **Grupo 1:** O Perfil Predominante (45% dos casos)

- Características: Homens, adultos jovens (~35 anos), Pardos, Solteiros.

## **Grupo 2:** Homens de Meia-Idade Estáveis (29.6% dos casos)

- Características: Homens, ~52 anos, Brancos, Casados.

## **Grupo 4:** O Perfil Feminino (20.8% dos casos)

- Características: Mulheres, ~38 anos, Brancas, Solteiras.

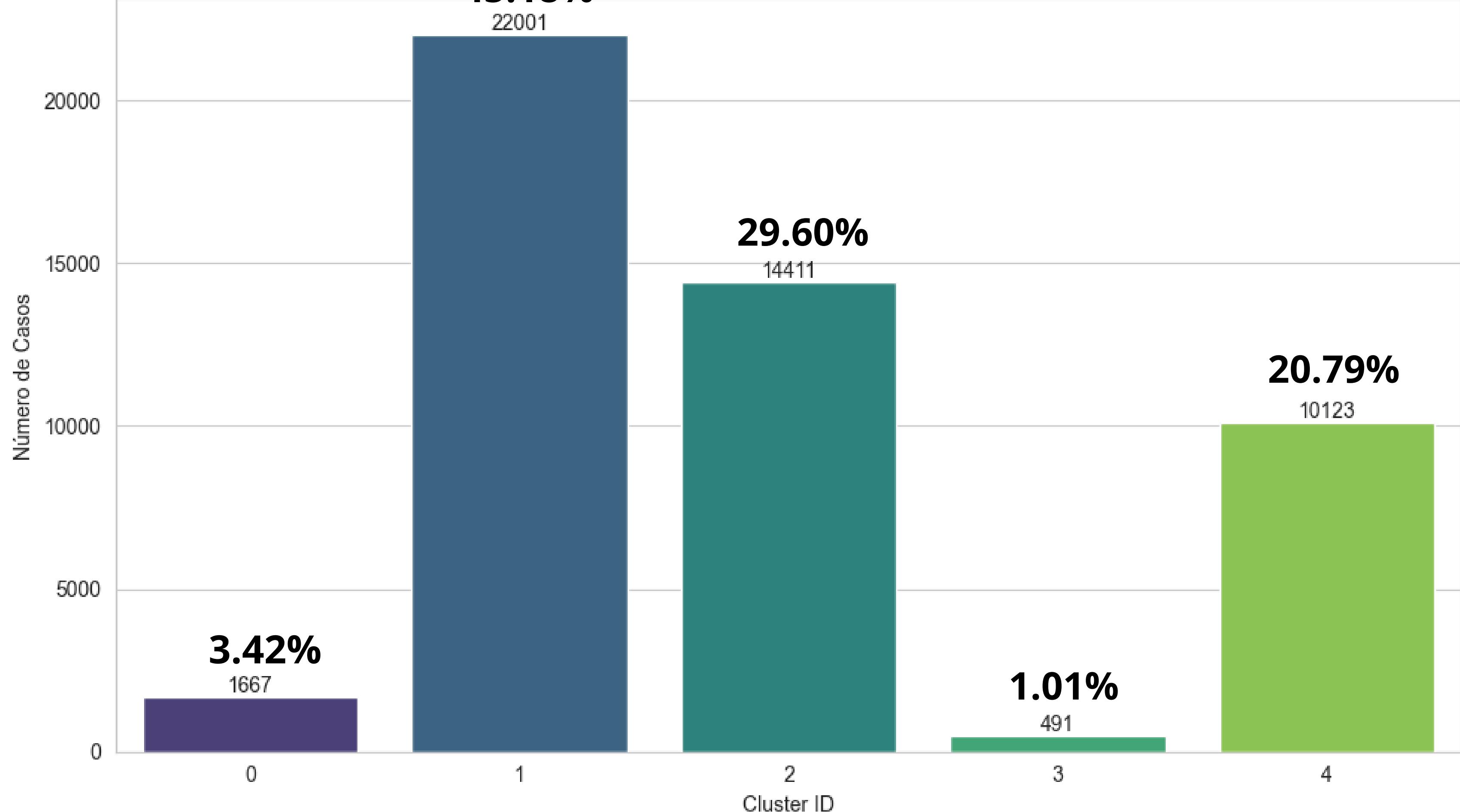
## **Grupo 0:** Idosos Viúvos (3.4% dos casos)

- Características: Homens, Idosos (~68 anos), Viúvos, Escolaridade mais baixa (4-7 anos).

## **Grupo 3:** Jovens Indígenas (1% dos casos)

- Características: Homens, Jovens (~25 anos), Indígenas.

## Tamanho dos Clusters Identificados



# **REGRAS DE ASSOCIAÇÃO**

# ATRIBUTOS SELECIONADOS PARA ANÁLISE

**Objetivo:** Identificar variáveis demográficas e clínicas relevantes para traçar perfis de mortalidade.

**Variáveis Mapeadas:**

**IDADE:** Transformada em faixas etárias de 15 em 15 anos (Ex: 0-14, 15-29, etc.) para reduzir a granularidade.

**CAUSABAS (Causa Básica):** A variável alvo da predição. Tratada como categoria textual.

**SEXO:** Variável categórica (Masculino/Feminino).

**RACACOR:** Identificação étnico-racial.

**ESTCIV (Estado Civil):** Solteiro, Casado, Viúvo, etc.

# PREPARAÇÃO DA BASE DE DADOS

## Estratégia de Limpeza e Transformação (ETL):

### Limpeza de Dados Faltantes:

Remoção de registos com excesso de valores nulos (threshold de 60% de colunas preenchidas) para garantir a qualidade das transações.

### Categorização (Feature Engineering):

Aplicação de prefixos identificadores para evitar ambiguidade entre números (Ex: SEXO\_1 vs ESC\_1).

Criação de "bins" (agrupamentos) para a idade, transformando um valor numérico contínuo em discreto.

### Criação de Transações:

Conversão do DataFrame tabular para um formato de lista de listas (transações).

# O ALGORITMO DE ASSOCIAÇÃO

**Metodologia:** Algoritmo Apriori

**Funcionamento:**

- Identifica conjuntos de itens frequentes (Frequent Itemsets) que aparecem juntos na base de dados.
- Gera regras do tipo "Se {A + B} então {C}".

**Parâmetros Configuradas no Código:**

**Suporte Mínimo (3%):** A regra deve aparecer em pelo menos 3% de todos os óbitos analisados. Garante relevância estatística.

**Confiança Mínima (35%):** Quando o perfil (Antecedente) ocorre, a Causa (Consequente) deve ocorrer em 35% das vezes.

**Lift Mínimo (1.2):** A associação deve ser pelo menos 20% mais forte do que o acaso.

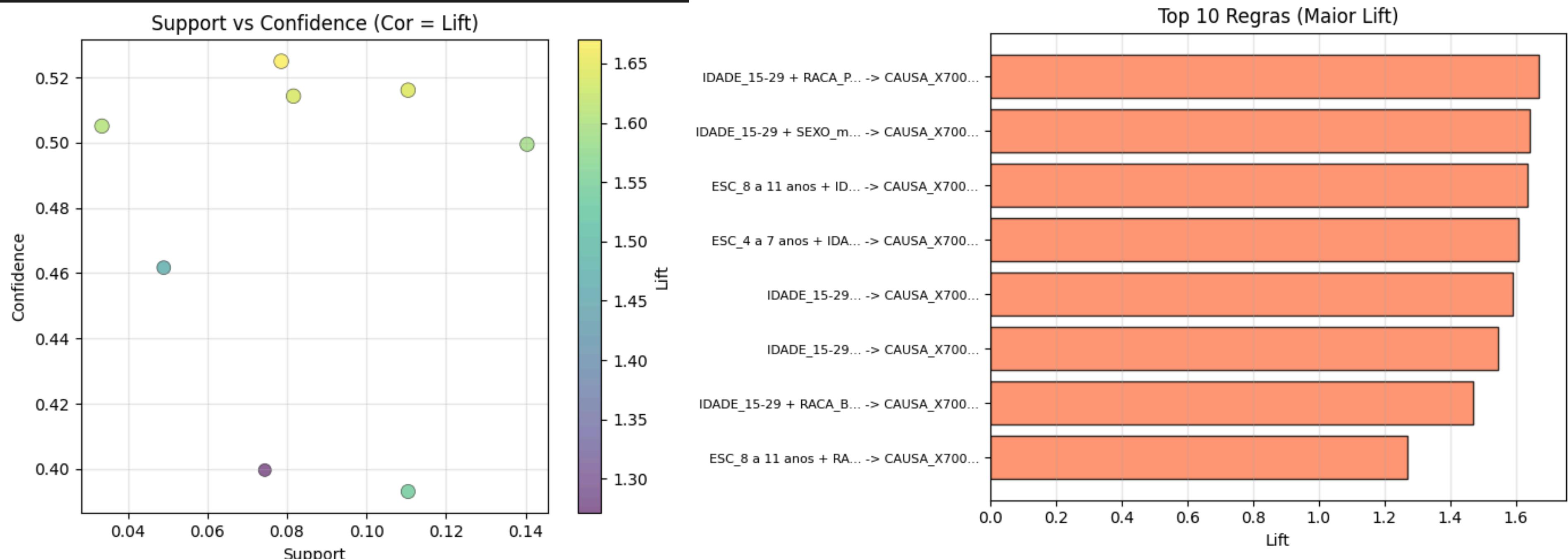
# ASSOCIAÇÕES IDENTIFICADAS

O Perfil Principal Os dados identificaram um grupo muito específico com forte tendência ao evento "Causa X700" (lesão autoprovocada) combinado com o estado civil Solteiro.

Quem faz parte desse grupo? O padrão é focado em Jovens (15 a 29 anos). A associação fica ainda mais forte quando esse jovem também possui uma dessas características adicionais:

- É de cor Parda;
- É do sexo Masculino;
- Tem escolaridade média (8 a 11 anos de estudo).

	<b>Antecedentes</b>	<b>Consequentes</b>	<b>support</b>	<b>confidence</b>	<b>lift</b>
555	IDADE_15-29 + RACA_Parda	CAUSA_X700 + ESTCIV_Solteiro	0.078574	0.524973	1.669878
563	IDADE_15-29 + SEXO_masculino	CAUSA_X700 + ESTCIV_Solteiro	0.110468	0.516120	1.641718
473	ESC_8 a 11 anos + IDADE_15-29	CAUSA_X700 + ESTCIV_Solteiro	0.081634	0.514297	1.635919
436	ESC_4 a 7 anos + IDADE_15-29	CAUSA_X700 + ESTCIV_Solteiro	0.033475	0.505113	1.606707
162	IDADE_15-29	CAUSA_X700 + ESTCIV_Solteiro	0.140390	0.499525	1.588932
565	IDADE_15-29	CAUSA_X700 + ESTCIV_Solteiro + SEXO_masculino	0.110468	0.393058	1.544603
550	IDADE_15-29 + RACA_Branca	CAUSA_X700 + ESTCIV_Solteiro	0.049021	0.461702	1.468622
486	ESC_8 a 11 anos + RACA_Parda	CAUSA_X700 + ESTCIV_Solteiro	0.074446	0.399625	1.271162



**OBRIGADO**