

Fundamentos de Machine Learning e Modelos Não-supervisionados [24E2_2]

Grupo 3

Diego dos Santos

Filipe Rodrigues

Leandro Ribeiro

Rafael Diniz

Business Understanding

O MJSP (Ministério da Justiça e Segurança Pública) está interessado em entender melhor seus dados criminais registrados pelo sistema SINESP (Sistema Nacional de Informações de Segurança Pública) através da Ciência de Dados.

Para isso, elaborou a seguinte questão a ser respondida:

Identificar padrões de crimes por estado para orientar políticas de segurança pública e alocação de recursos.

Referências Bibliográficas

Ministério da Justiça e Segurança Pública <https://www.gov.br/mj/pt-br> SINESP
<https://seguranca.sinesp.gov.br/sinesp-seguranca>

Base Populacional por UF do IBGE.

<https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html>

Data Understanding

Os dados disponíveis para esta análise foram extraídos das soluções SinespJC e Sinesp Integração, fontes primárias dos seguintes

Indicadores:

- Totais de Vítimas de homicídio doloso, lesão corporal seguida de morte, roubo seguido de morte.
- Cada linha da base de dados representa em um mês o número de vítimas de um Estado.
- Total de linhas: 19748
- Período: 2015 - 2022

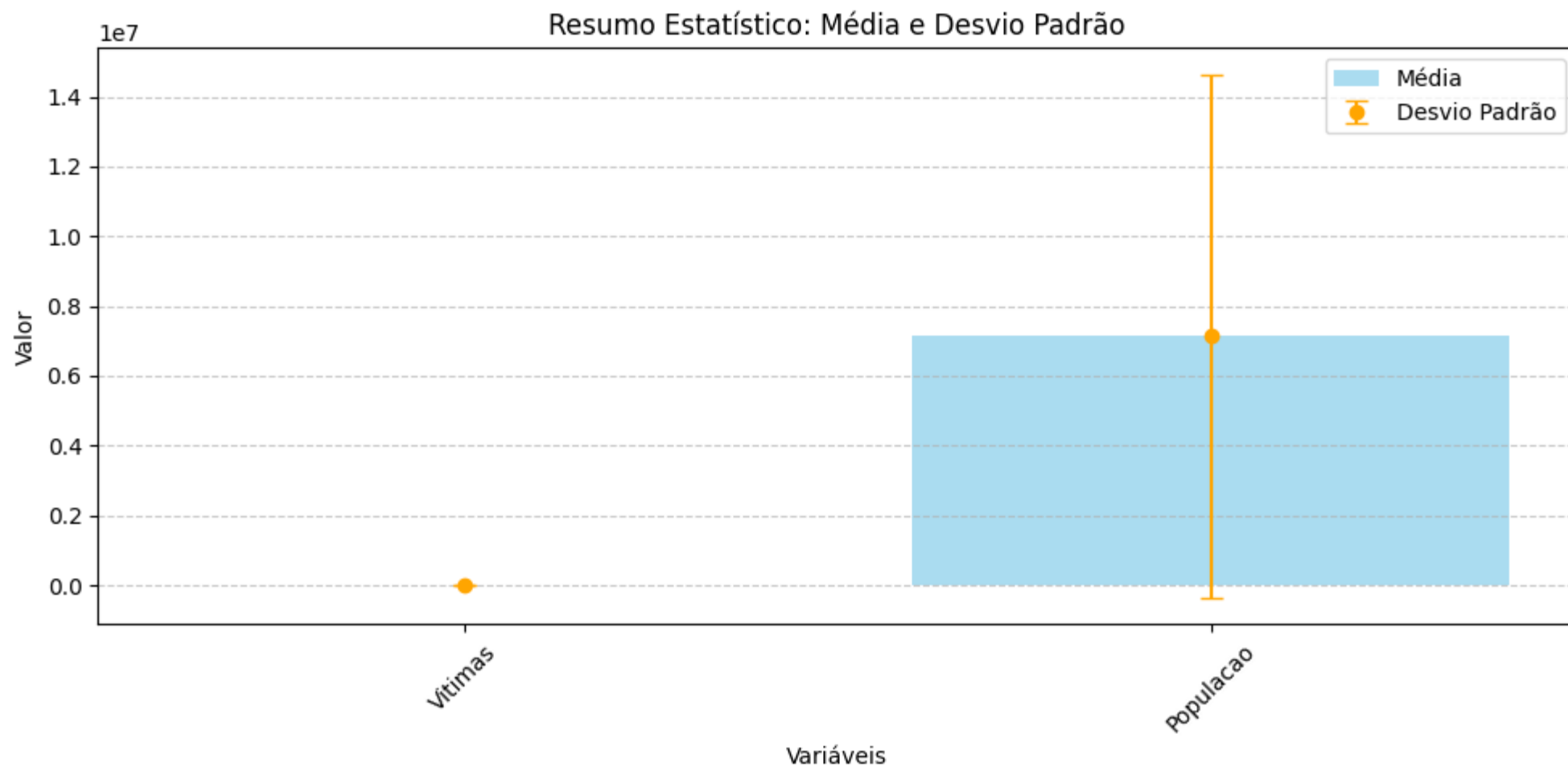
Variáveis Aleatórias

Colunas	Tipo
UF	String
Tipo_Crime	String
Ano	Inteiro
Mes	String
Sexo	String
Vitimas	Inteiro
Populacao	Inteiro

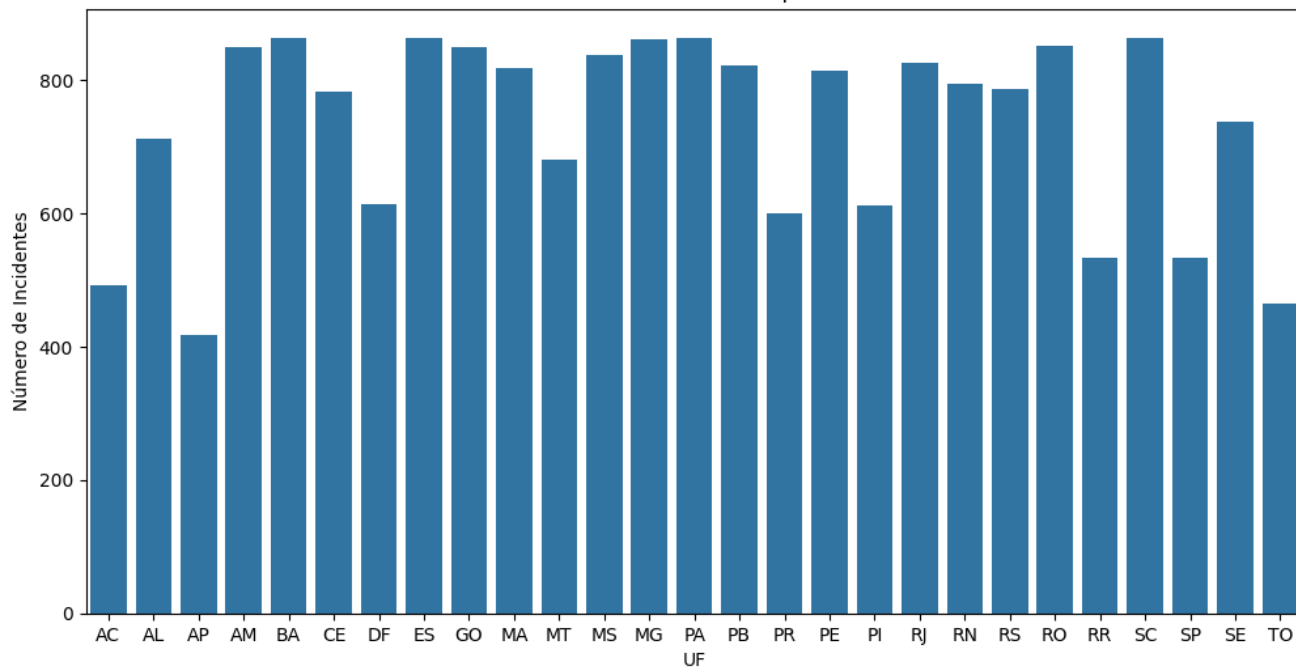
Ações de Limpeza:

- Padronizar dados em caixa alta
- Remover acentos, caracteres especiais e espaço
- Sem outliers

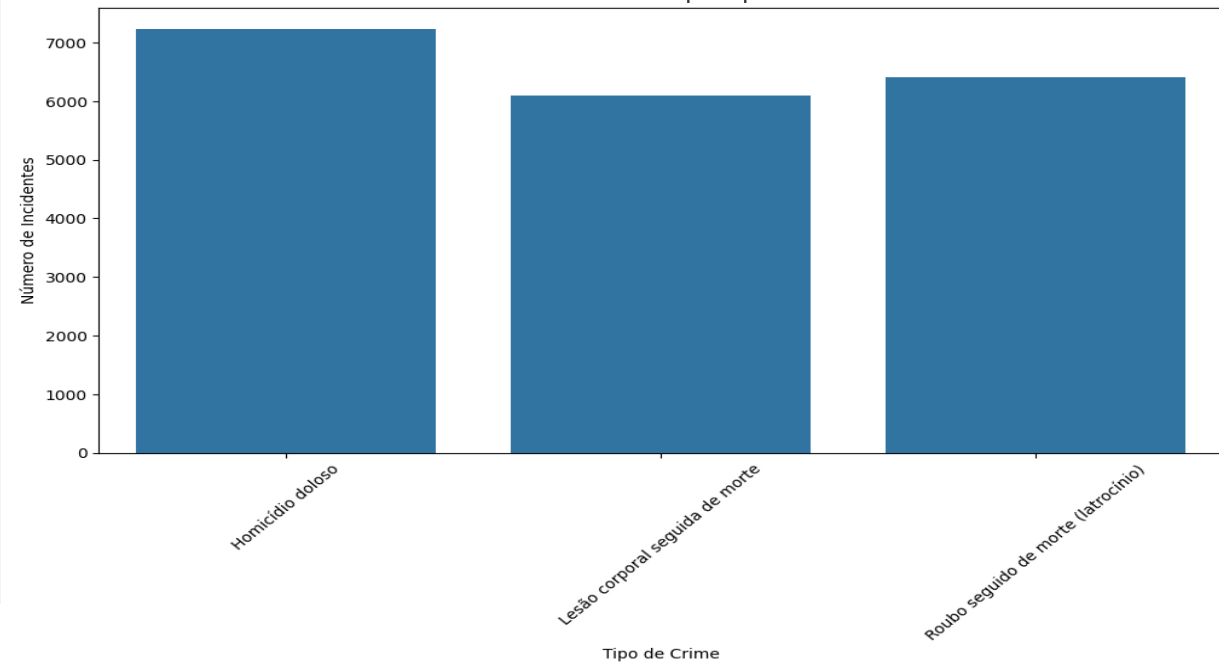
Análise Exploratória



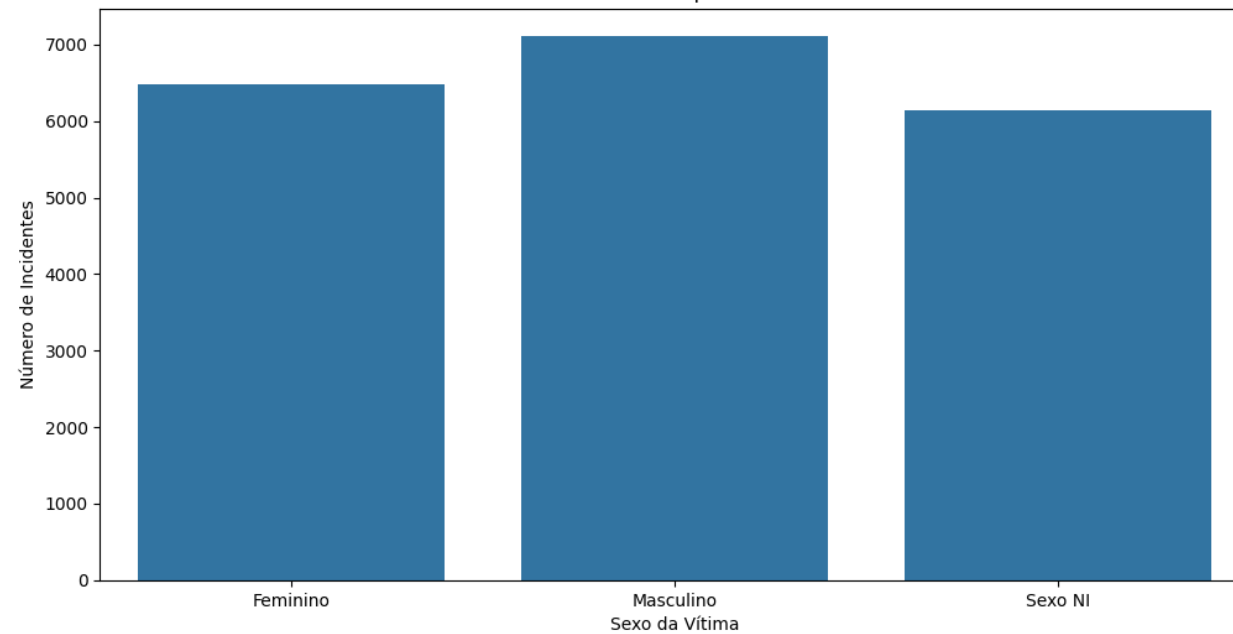
Número de Incidentes por UF



Número de Incidentes por Tipo de Crime



Número de Incidentes por Sexo da Vítima



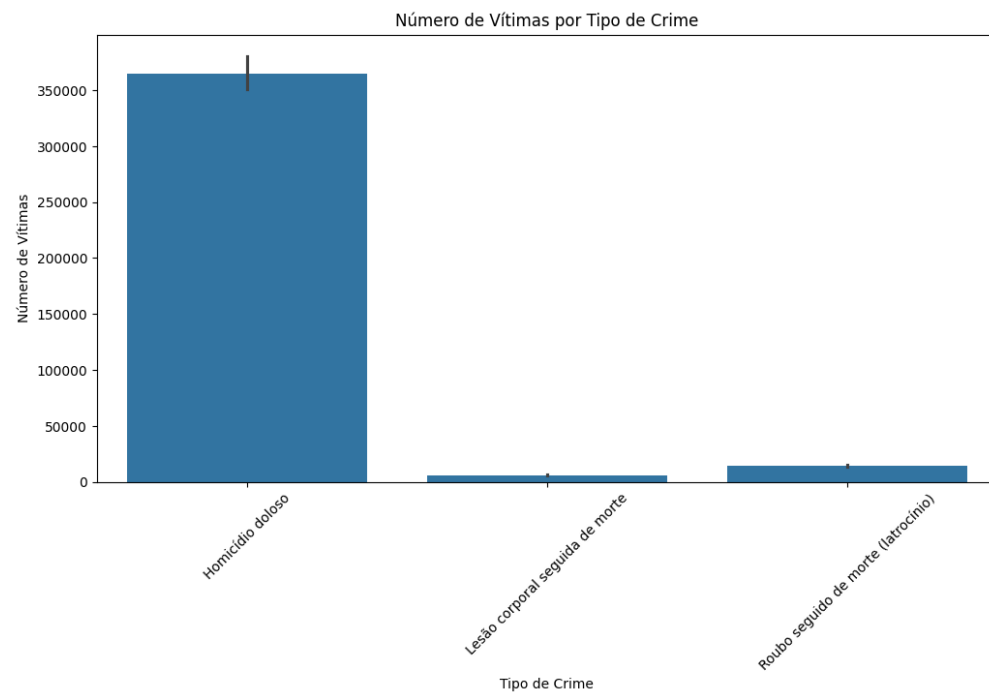
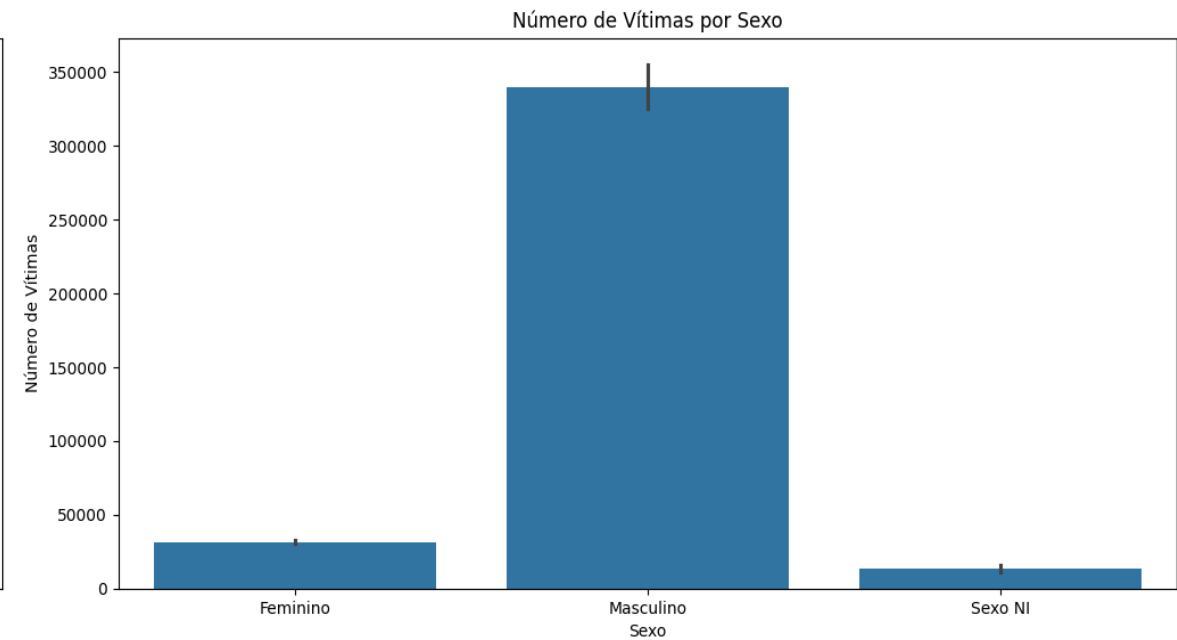
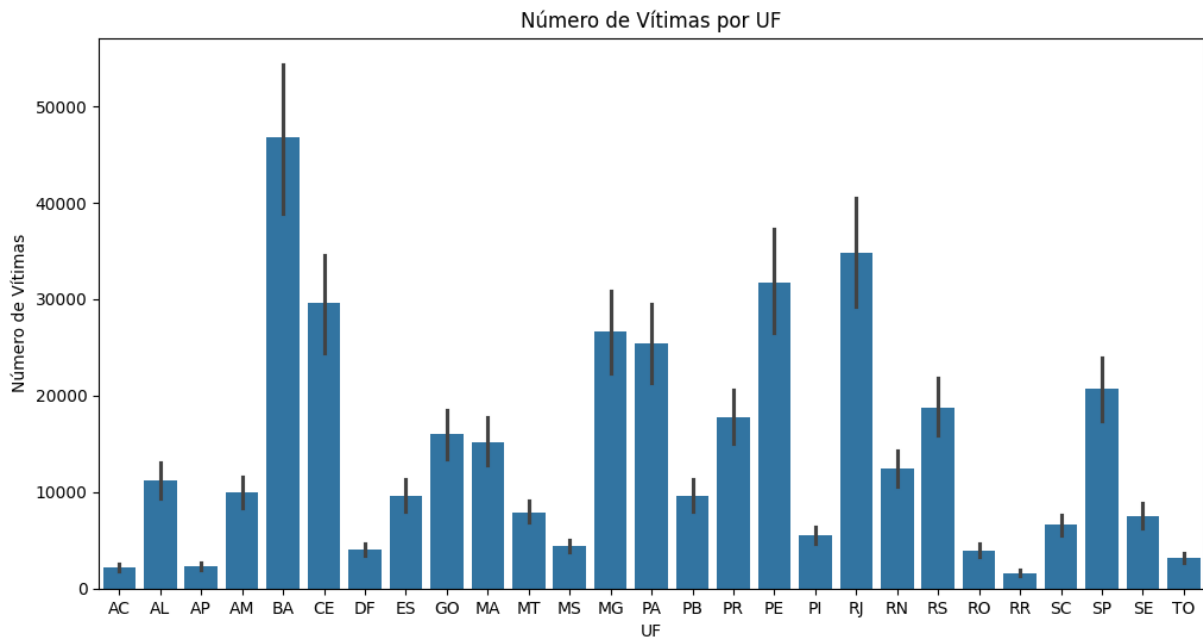


Tabela de Contingência: UF vs Tipo de Crime

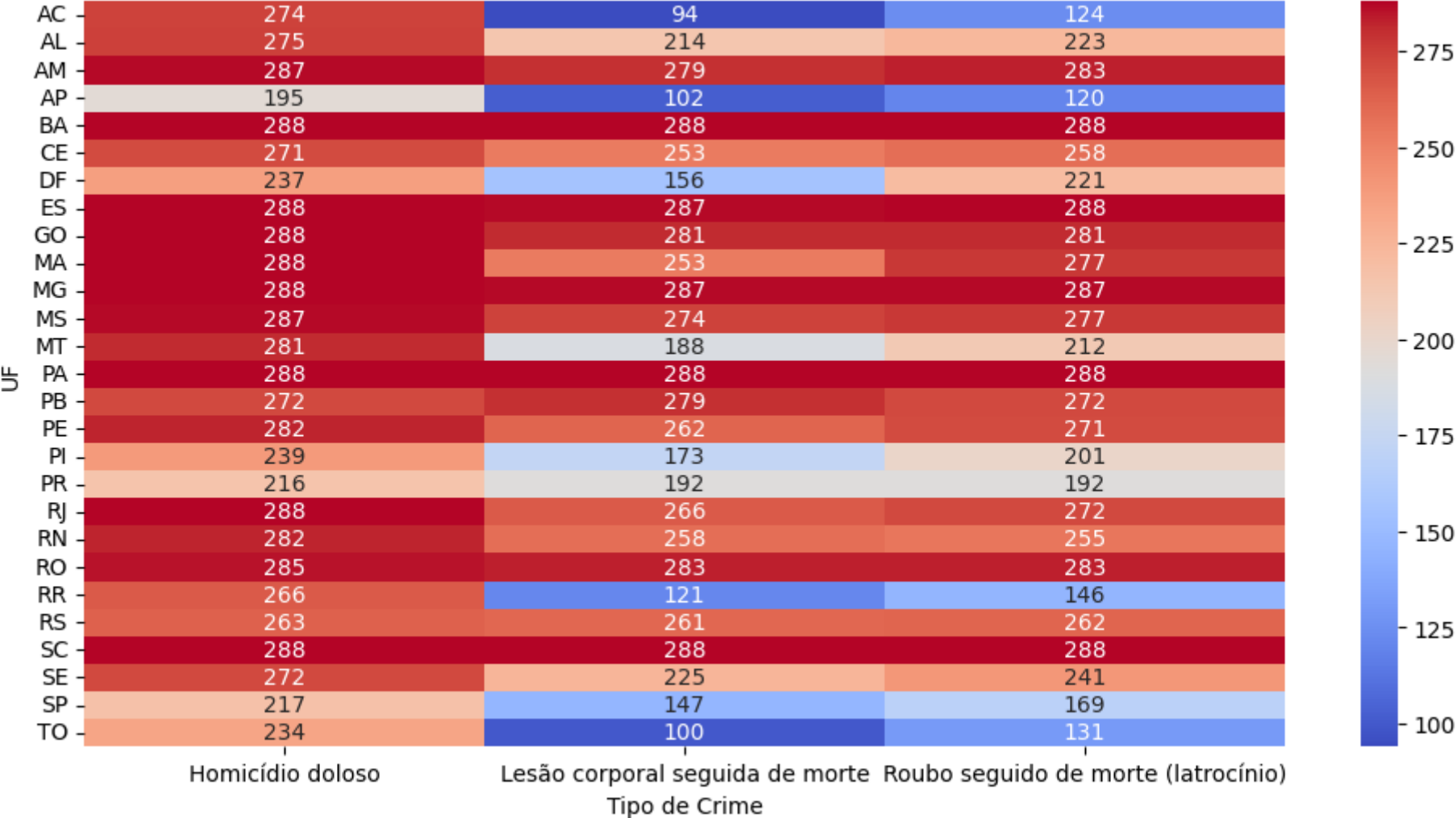
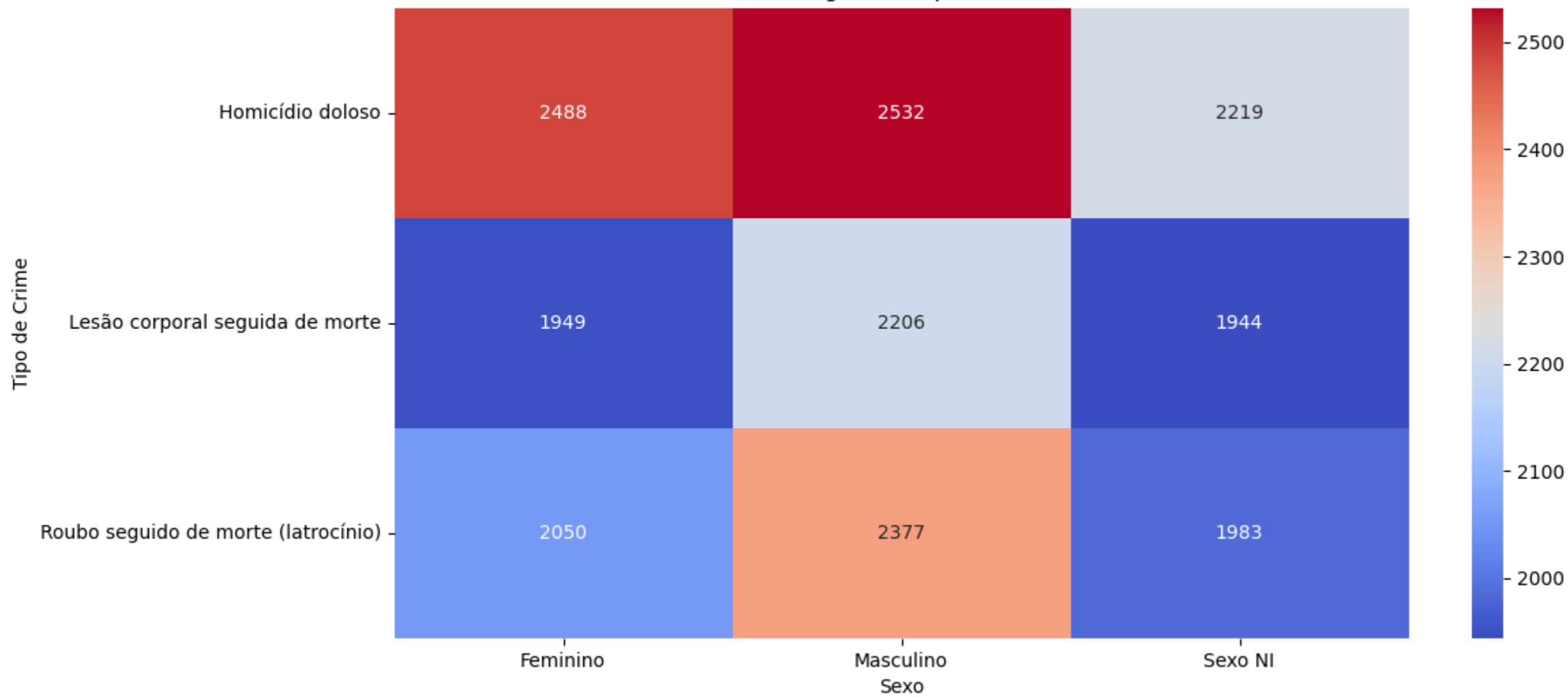
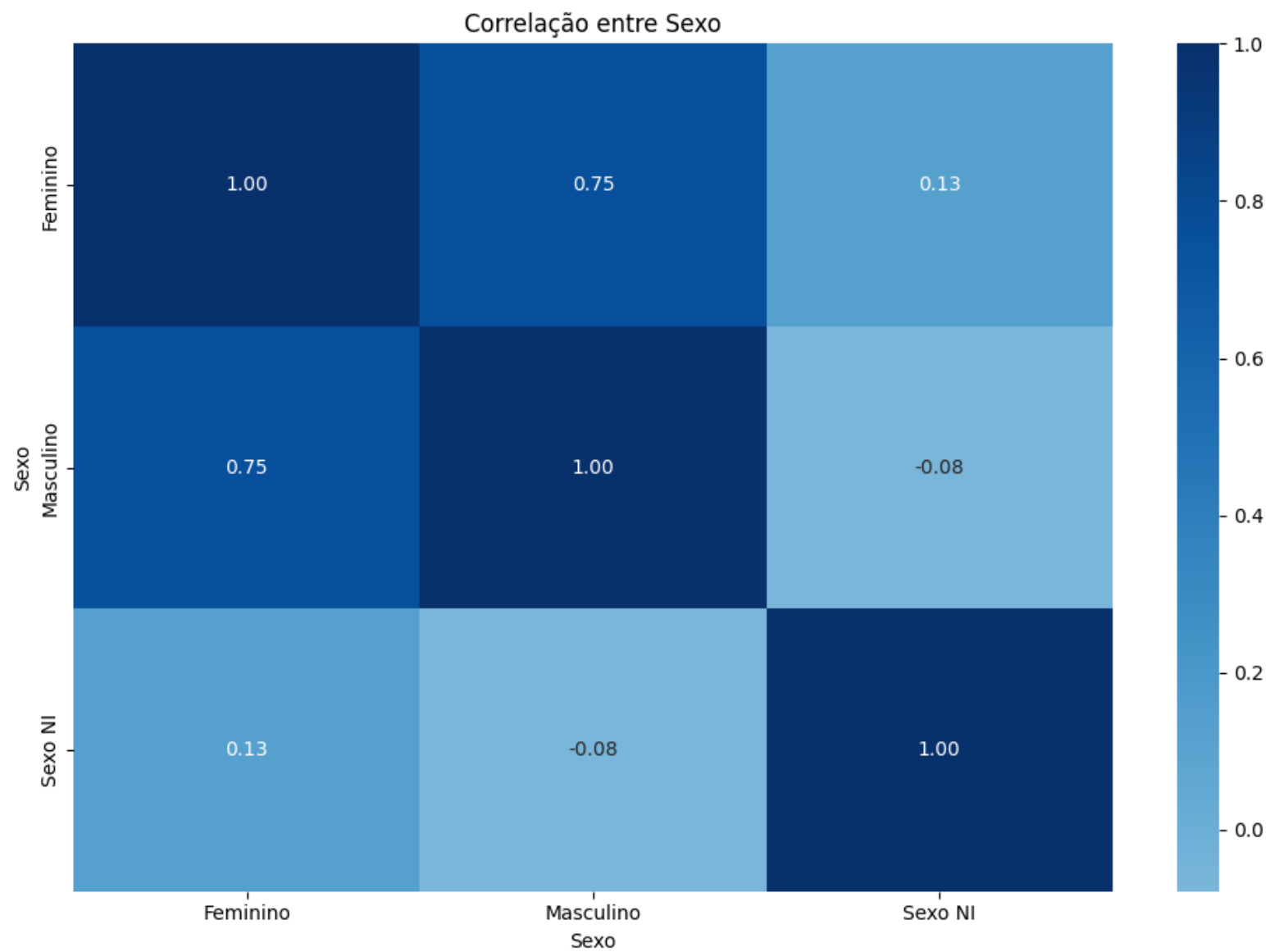
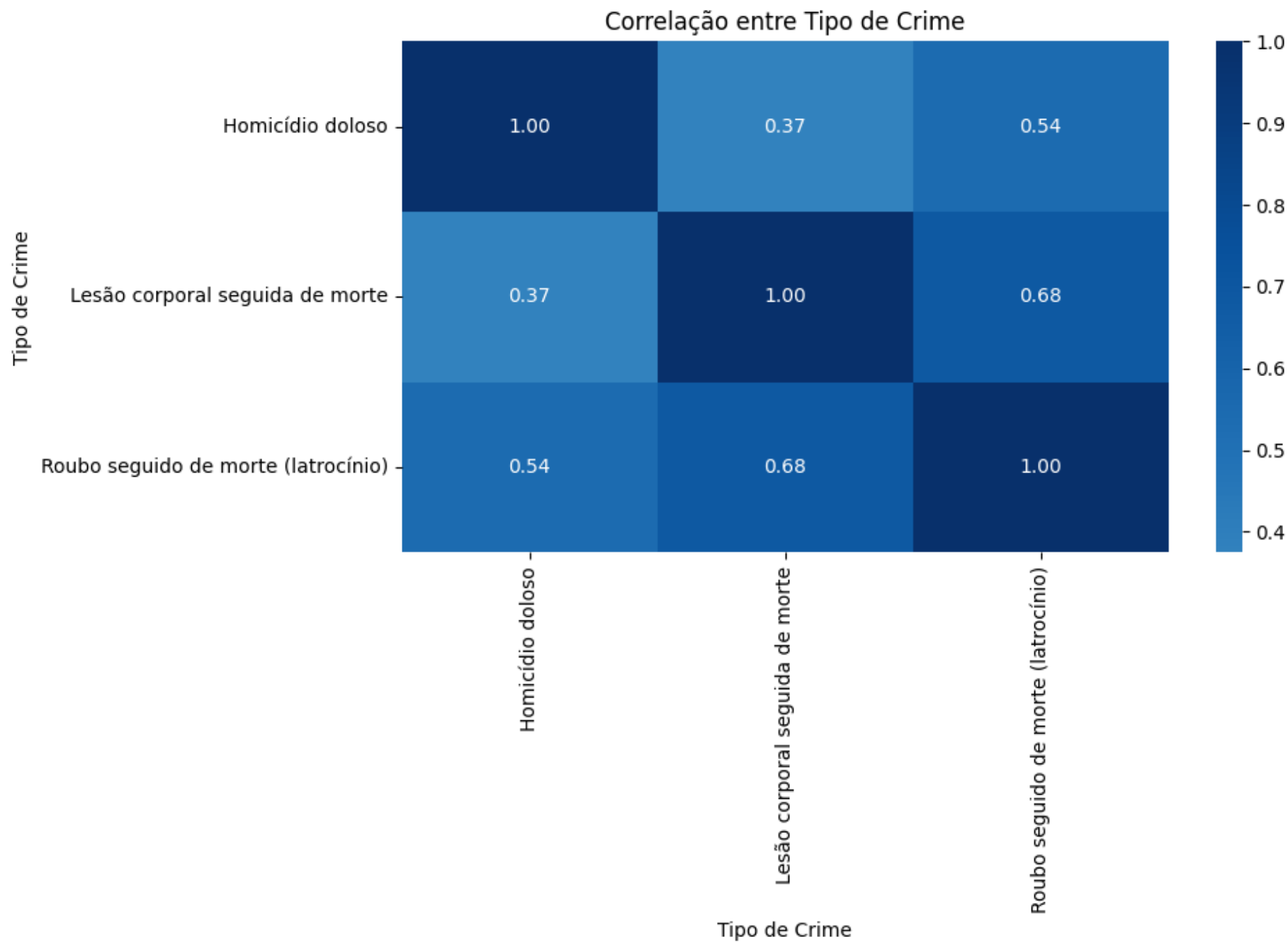


Tabela de Contingência: Tipo de Crime vs Sexo







Insights Obtidos

- Insights da Base de Dados:

1. Disparidade de Gênero na Criminalidade:

- Homens são mais frequentemente alvo de crimes em todos os estados, sugerindo uma disparidade de gênero nos índices de criminalidade.

2. Impacto da Política Pública na Segurança:

- Estados com menor investimento em políticas públicas de segurança apresentam números mais elevados de vítimas de crimes, indicando desafios adicionais no combate à criminalidade devido à falta de recursos.

3. Correlação entre População e Criminalidade:

- Estados mais populosos tendem a ter um número absoluto maior de vítimas por mês, refletindo a relação entre densidade populacional e incidência de crimes. Esses insights podem subsidiar políticas de segurança pública direcionadas para áreas com maior densidade populacional.

Data Preparation

- Quantificação dos Atributos: Os atributos são quantificados principalmente através da contagem de vítimas de diferentes tipos de crimes em cada estado, mês e ano.
- Atributos Considerados:
 - 1. UF (Estado): Essencial para identificar a localização geográfica dos dados.
 - 2. Tipo de Crime: Indica o tipo de crime ocorrido.
 - 3. Ano e Mês: Permitem a análise temporal dos dados, identificando padrões ao longo do tempo.
 - 4. Vítimas: O número de vítimas de cada tipo de crime é o principal atributo de interesse neste contexto, pois é o que está sendo analisado e comparado.
- A escolha desses atributos é justificada pela necessidade de entender a distribuição e o impacto dos diferentes tipos de crimes em cada estado ao longo do tempo, levando em consideração as variações na população.
- Transformações: Esses dados são agregados e normalizados pela população, o que permite uma análise comparativa entre os estados, levando em consideração o tamanho da população de cada um.

Modeling

Modelo de Clusterização Hierárquica:

- Algoritmo: Utilização da clusterização hierárquica devido à capacidade de lidar com estruturas hierárquicas nos dados, capturando padrões em diferentes níveis de granularidade.
- Hiperparâmetro: Escolha do método de ligação "average" para calcular as distâncias entre clusters, levando em consideração a média das distâncias entre todos os pares de pontos nos dois grupos, robustecendo a análise contra outliers e ruídos.
- Figura de Mérito: Avaliação da qualidade dos clusters através do Índice de Silhueta, que quantifica a coesão intra-cluster e a separação inter-cluster, fornecendo uma medida objetiva da eficácia do método de ligação "average" na definição dos clusters.

Índice de Silhueta para average: 0.5807549211954753

Melhor método baseado no Índice de Silhueta: average com score 0.5807549211954753

Evaluation

Cluster A (Verde):

- Estados: RS, MS, PA, PR, SP, SC.
- Características dos Crimes:
 - Homicídio doloso: Média de 48,12 vítimas
 - Lesão corporal seguida de morte: Média de 1,02 vítimas
 - Roubo seguido de morte (latrocínio): Média de 2,41 vítimas
- Características por Sexo:
 - Feminino: Média de 5,65 vítimas
 - Masculino: Média de 41,49 vítimas
 - Sexo NI (não informado): Média de 5,00 vítimas

- Cluster B (Laranja):

- Estados: ES, AM, AC, RO, MT, PB, GO, TO, MA, PI, PE, AL, SE, BA, CE.
- Características dos Crimes:
 - Homicídio doloso: Média de 41,70 vítimas
 - Lesão corporal seguida de morte: Média de 0,63 vítimas
 - Roubo seguido de morte (latrocínio): Média de 1,80 vítimas
- Características por Sexo:
 - Feminino: Média de 3,82 vítimas
 - Masculino: Média de 39,31 vítimas
 - Sexo NI (não informado): Média de 1,60 vítimas

Evaluation

- - Cluster C (Azul):
- Estados: MG, DF.
- Características dos Crimes:
 - Homicídio doloso: Média de 60,38 vítimas
 - Lesão corporal seguida de morte: Média de 1,36 vítimas
 - Roubo seguido de morte (latrocínio): Média de 2,44 vítimas
- Características por Sexo:
 - Feminino: Média de 5,29 vítimas
 - Masculino: Média de 60,97 vítimas
 - Sexo NI (não informado): Média de 0,57 vítimas

Cluster	Sexo	Vitimas
1	Feminino	5,65
1	Masculino	41,49
1	Sexo NI	5,00
2	Feminino	3,82
2	Masculino	39,31
2	Sexo NI	1,60
3	Feminino	5,29
3	Masculino	60,97
3	Sexo NI	0,57

Cluster	Tipo_Crime	Vitimas
1	Homicídio doloso	48,12
1	Lesão corporal seguida de morte	1,02
1	Roubo seguido de morte (latrocínio)	2,41
2	Homicídio doloso	41,70
2	Lesão corporal seguida de morte	0,63
2	Roubo seguido de morte (latrocínio)	1,80
3	Homicídio doloso	60,38
3	Lesão corporal seguida de morte	1,36
3	Roubo seguido de morte (latrocínio)	2,44

Evaluation

- **Cluster Verde**

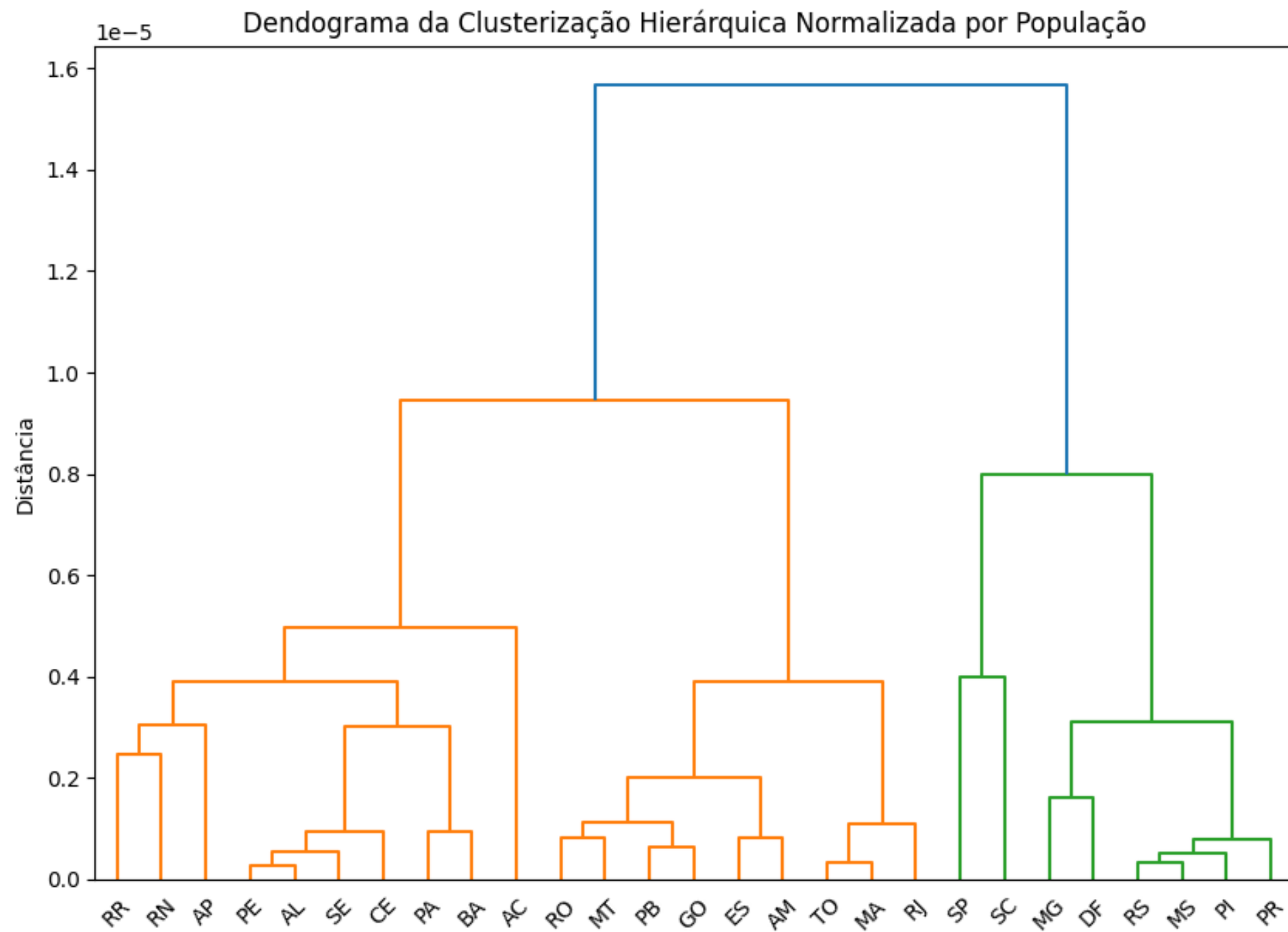
Este cluster possui uma alta média de homicídios dolosos e latrocínios. As vítimas são predominantemente masculinas, mas também há uma quantidade significativa de vítimas femininas.

- **Cluster Laranja**

Este cluster apresenta uma média ligeiramente menor de homicídios dolosos em comparação com o Cluster 1, mas ainda significativa. Novamente, as vítimas masculinas são predominantes

- **Cluster Azul**

Este cluster tem a maior média de homicídios dolosos entre os três clusters. A média de vítimas masculinas é a mais alta, indicando uma alta prevalência de violência letal.



Próximos Passos

1. Políticas de Segurança Personalizadas:

Desenvolver programas de prevenção da violência, incluindo iniciativas comunitárias e educativas, para reduzir o número de homicídios dolosos.

2. Recursos Alocados de Forma Otimizada:

Direcionar mais recursos de segurança para estados no Cluster Azul, que enfrentam desafios maiores de violência. Isso pode incluir investimentos em policiamento, programas sociais e infraestrutura de segurança.

3. Estudos Adicionais:

Investigar as causas subjacentes dos crimes, incluindo fatores socioeconômicos e demográficos. Realizar estudos para entender melhor como a pobreza, educação e outros fatores influenciam os padrões de criminalidade.

Conclusão

A análise do dendrograma com base nos três tipos de crimes disponíveis revela padrões distintos de criminalidade entre os estados brasileiros, permitindo uma abordagem mais direcionada e eficaz na implementação de políticas de segurança pública. A personalização das estratégias de combate ao crime, a alocação otimizada de recursos e a análise contínua são fundamentais para melhorar a segurança e reduzir a incidência de crimes em todo o país.