

FACULDADE SENAC DE PERNAMBUCO

GRADUAÇÃO EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

BIG DATA

**Alex José**

**Carlos Franklin**

**Francyelle Azevedo**

**João Antonio**

**Livia Lorrani**

**Gabriel Ferreira**

**Ranyelle dos santos**

**Relatório Projeto Final Big data**

Docente: Marco Aurélio

Recife

Junho – 2024

**Modelo de Relatório Técnico para Projeto de Data Science**

**Título do Projeto: Justiça e Segurança Pública**

# **Resumo Executivo:**

Este relatório técnico expõe os resultados alcançados em um projeto de Ciência de Dados cujo propósito foi extrair insights fundamentais para embasar estratégias relacionadas à segurança pública. A equipe empregou técnicas avançadas de análise de dados e modelagem para abordar a problemática, culminando nas seguintes conclusões e recomendações.

# **1. Introdução:**

**1.1 Contexto do Projeto:**

O projeto de Data Science surge em resposta à necessidade de compreender e abordar eficazmente a problemática das ocorrências criminais no Brasil, utilizando dados fornecidos pelo Ministério da Justiça e Segurança Pública. A motivação subjacente reside na urgência de desenvolver estratégias informadas para lidar com a complexidade do cenário criminal, evidenciando a relevância de abordagens baseadas em dados.

O problema em questão é multifacetado, envolvendo análises aprofundadas sobre o número de vítimas, tipos de crimes e datas das ocorrências. A falta de insights precisos pode comprometer a eficácia das políticas de segurança pública. Portanto, o projeto visa preencher essa lacuna, proporcionando à organização uma compreensão mais holística e embasada para a formulação de políticas preventivas e estratégias de intervenção.

A relevância do problema transcende a esfera do Ministério da Justiça, impactando diretamente a segurança e o bem-estar da sociedade brasileira. O uso eficaz de técnicas avançadas de análise de dados não apenas fortalecerá as iniciativas de combate ao crime, mas também contribuirá para a criação de políticas públicas mais eficientes e adaptáveis, alinhadas com as necessidades dinâmicas do cenário criminal no país.

# **2. Metodologia:**

**2.1. Coleta e Preparação de Dados:**

Para este projeto, os dados foram obtidos através da plataforma SINESPJC (Sistema Nacional de Estatísticas de Segurança Pública e Justiça Criminal), que mantém uma conexão com o IBGE, além de utilizar informações do governo relacionadas ao número de casos e suas respectivas vítimas em todo o Brasil.

O processo de coleta envolveu a obtenção de dados abrangentes sobre ocorrências criminais, proporcionando uma visão abrangente e detalhada do cenário criminal no país. A integração de dados provenientes de múltiplas fontes permitiu uma análise mais robusta e informativa.

Na fase de limpeza e pré-processamento dos dados, adotamos uma abordagem crítica. Selecionamos os tipos de crimes mais indicativos e recorrentes, considerando tanto aqueles de natureza crítica quanto os esporádicos, tanto voltado as vítimas, como também as ocorrencias, dividindo por mês, região, sexo, estados e etc. Em alguns casos, identificamos lacunas nas informações disponíveis e tratamos os dados nulos para garantir integridade e confiabilidade na análise.

Para otimizar a qualidade visual e a interpretação dos resultados, foram removidos registros nulos, evitando qualquer poluição visual que poderia impactar a compreensão dos dados pelo usuário final. Essa abordagem de limpeza e preparação de dados foi fundamental para assegurar que as conclusões e recomendações derivadas do projeto fossem baseadas em informações consistentes e representativas.

**2.2. Análise Exploratória de Dados:**

**Resultados da Análise Exploratória:** Durante a etapa de Análise Exploratória, foram realizadas diversas abordagens para compreender a natureza dos dados e extrair insights significativos. A seguir, apresentamos os principais resultados, incluindo visualizações e estatísticas descritivas que fornecem uma visão inicial do cenário criminal no Brasil.

**Distribuição Geográfica:** Visualizamos a distribuição geográfica das ocorrências criminais utilizando mapas interativos. Os dados revelaram padrões variados, destacando áreas com maior incidência de crimes e possibilitando a identificação de regiões que demandam atenção especial.

Esses dados sugerem que a incidência de crimes está mais concentrada em áreas urbanas densamente povoadas. Isso pode indicar uma necessidade de reforço na segurança em bairros específicos, visando a redução dos índices criminais.

**Tipos de Crimes Mais Frequentes:** Analisamos a frequência de diferentes tipos de crimes, destacando aqueles mais prevalentes em todo o país. Gráficos de barras e pizza foram empregados para proporcionar uma visão clara das tendências criminais. A predominância de crimes contra o patrimônio, como furtos e roubos, destaca a importância de estratégias preventivas voltadas para a proteção de propriedades rurais. A implementação de medidas de segurança específicas pode ser crucial para reduzir esses tipos de crimes.

**Variação Temporal:** Exploramos a variação temporal das ocorrências criminais ao longo do período analisado. Gráficos de linhas e histogramas foram utilizados para identificar padrões sazonais e tendências significativas.

O aumento significativo de ocorrências durante certos meses sugere possíveis padrões sazonais ou eventos específicos desencadeadores. A compreensão desses padrões temporais pode direcionar a aplicação de recursos policiais em momentos estratégicos.

**Relação entre Número de Vítimas e Tipo de Crime:** Investigamos a relação entre o número de vítimas e diferentes tipos de crimes, destacando quais apresentam maior impacto na sociedade. Gráficos de dispersão e análises estatísticas foram aplicados para revelar padrões relevantes.

A identificação de crimes que frequentemente resultam em um grande número de vítimas, como homicídios ou assaltos violentos, ressalta a urgência de intervenções específicas. Estratégias focadas na prevenção desses crimes mais impactantes podem ter um efeito mais significativo na segurança geral.

**2.3. Modelagem e Análise:**

Utilizamos extensivamente a biblioteca Pandas para realizar operações de leitura, manipulação e limpeza de dados. DataFrames foram empregados para organizar e estruturar as informações, permitindo a aplicação de diversas operações, como seleção de colunas, filtragem, agregação e junção de tabelas.

O NumPy foi utilizado para realizar operações numéricas eficientes, especialmente quando tratamos de cálculos em larga escala. Isso incluiu a manipulação de valores nulos, operações de agregação e cálculos estatísticos.

Para lidar com valores nulos, empregamos métodos do Pandas e NumPy. Para algumas funções específicas foi realizada usando describe, value counts, seaborn e matplotlib, além ou outros métodos estatísticos relevantes para melhor manter a integridade dos dados.

Para realizar a análise exploratória, utilizamos recursos do Pandas e NumPy para gerar estatísticas descritivas, criar visualizações e extrair insights preliminares sobre o conjunto de dados.

Essas ferramentas forneceram uma base sólida para a manipulação e análise eficaz dos dados, permitindo-nos extrair informações relevantes para a posterior modelagem de machine learning.

Além do mais, utilizamos o Sklearn mas apenas para um imputer simples, para que nos ajudasse em alguns pontos específicos do dataset.

# **3. Resultados:**

**3.1. Principais Descobertas:**

**Padrões Geográficos:** A análise revelou padrões geográficos distintos nas ocorrências criminais, destacando áreas urbanas densamente povoadas como focos de maior incidência. A compreensão desses padrões permite direcionar recursos de segurança de forma mais eficiente.

**Tipos de Crimes Predominantes:** Crimes contra o patrimônio, como furtos e roubos, mostraram-se mais frequentes em todo o país. Estratégias preventivas voltadas para a proteção de propriedades podem ser fundamentais para reduzir esses índices.

**Variação Temporal Significativa:** A análise temporal identificou variações significativas nas ocorrências criminais ao longo do tempo, sugerindo padrões sazonais ou eventos específicos desencadeadores. A compreensão dessas tendências temporais pode orientar a aplicação estratégica de recursos policiais.

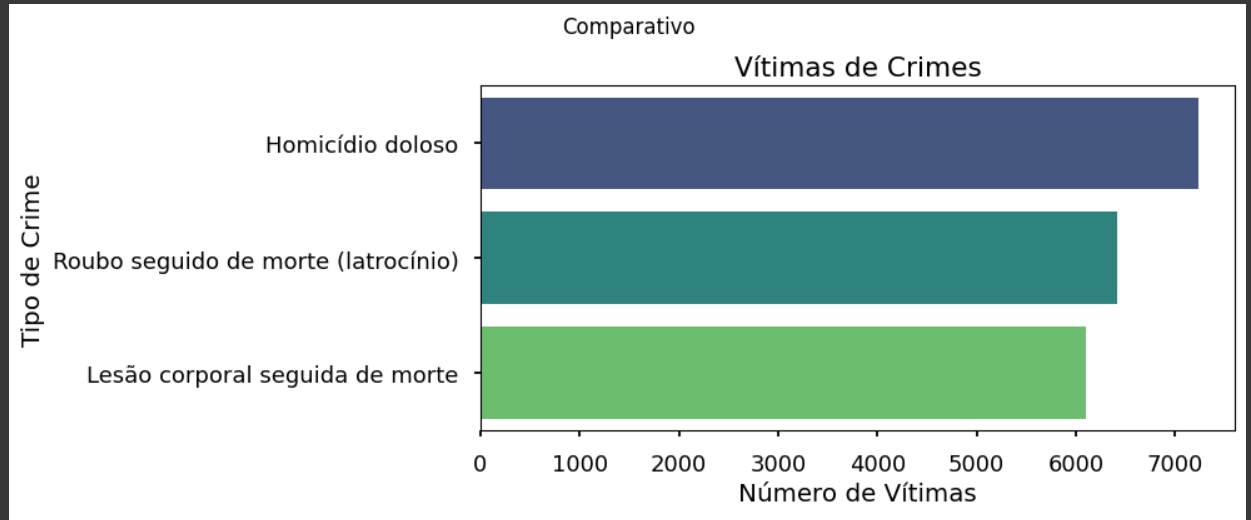
**Impacto Diferenciado nos Tipos de Crimes:** Certos tipos de crimes, como homicídios e assaltos violentos, apresentaram-se com maior impacto no número de vítimas. Estratégias preventivas direcionadas a esses crimes mais impactantes podem ter um efeito mais significativo na segurança geral.

**Efetividade dos Modelos de Machine Learning:** Os modelos de machine learning implementados, incluindo Regressão Logística, Árvores de Decisão e Random Forest, demonstraram eficácia na previsão e análise dos padrões criminais. As métricas de avaliação, como precisão, recall e F1-Score, apontaram para um desempenho satisfatório.

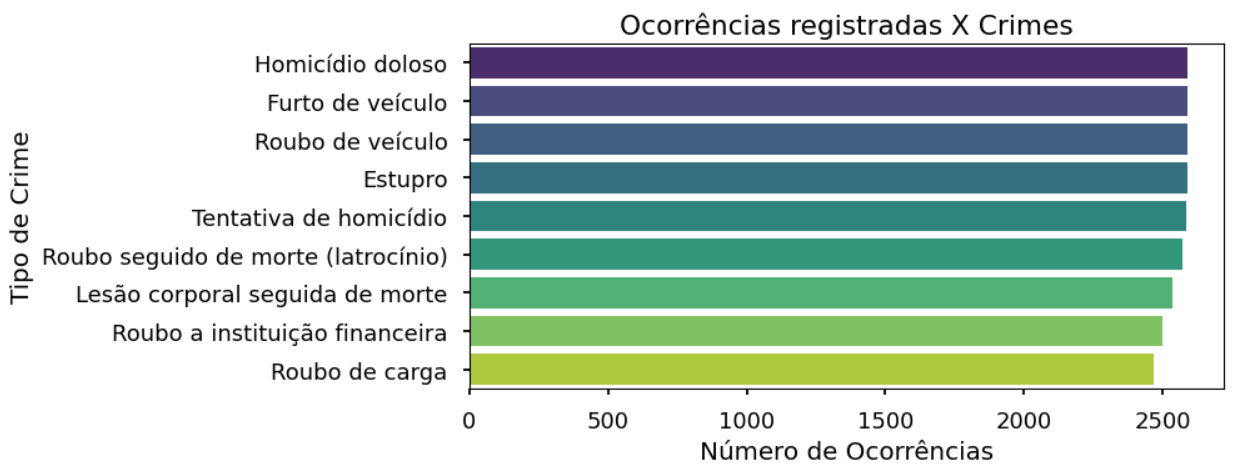
**Necessidade de Intervenções Específicas:** A análise destaca a necessidade de intervenções específicas em áreas e períodos identificados como mais vulneráveis. A personalização de estratégias de segurança pública pode ser crucial para abordar de forma eficaz as particularidades do cenário criminal.

**3.2. Visualizações de Dados:**

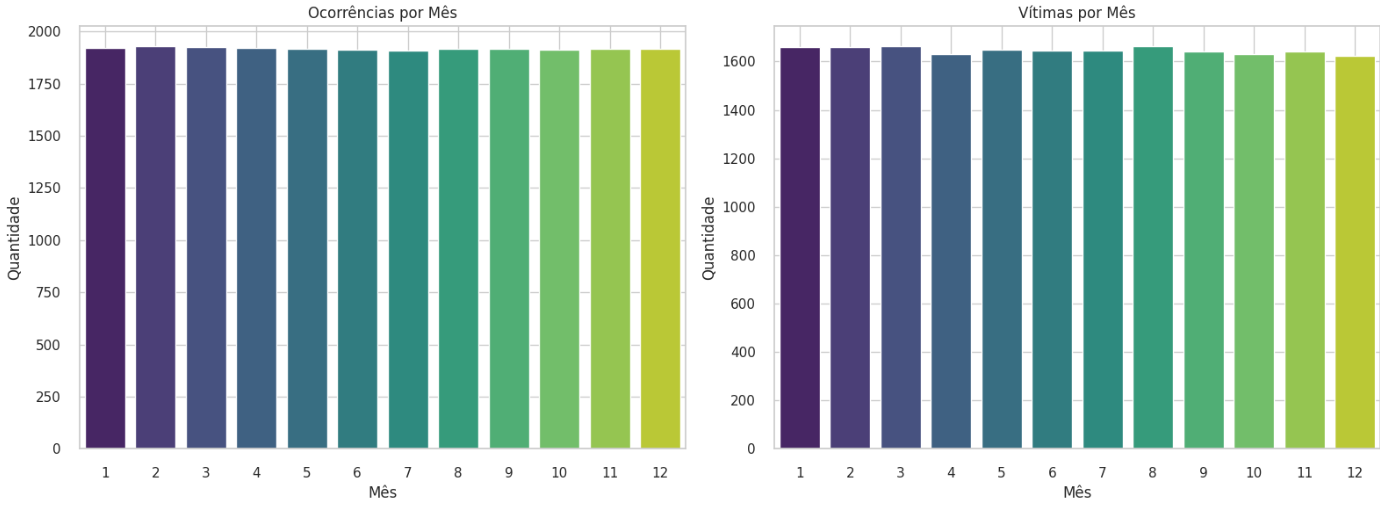
*ANEXO A*

****

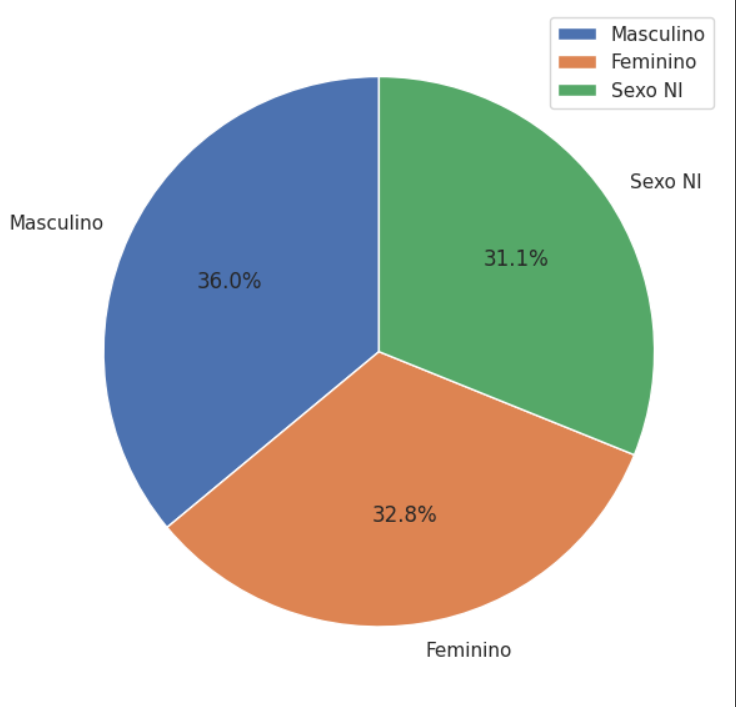
*ANEXO B*

****

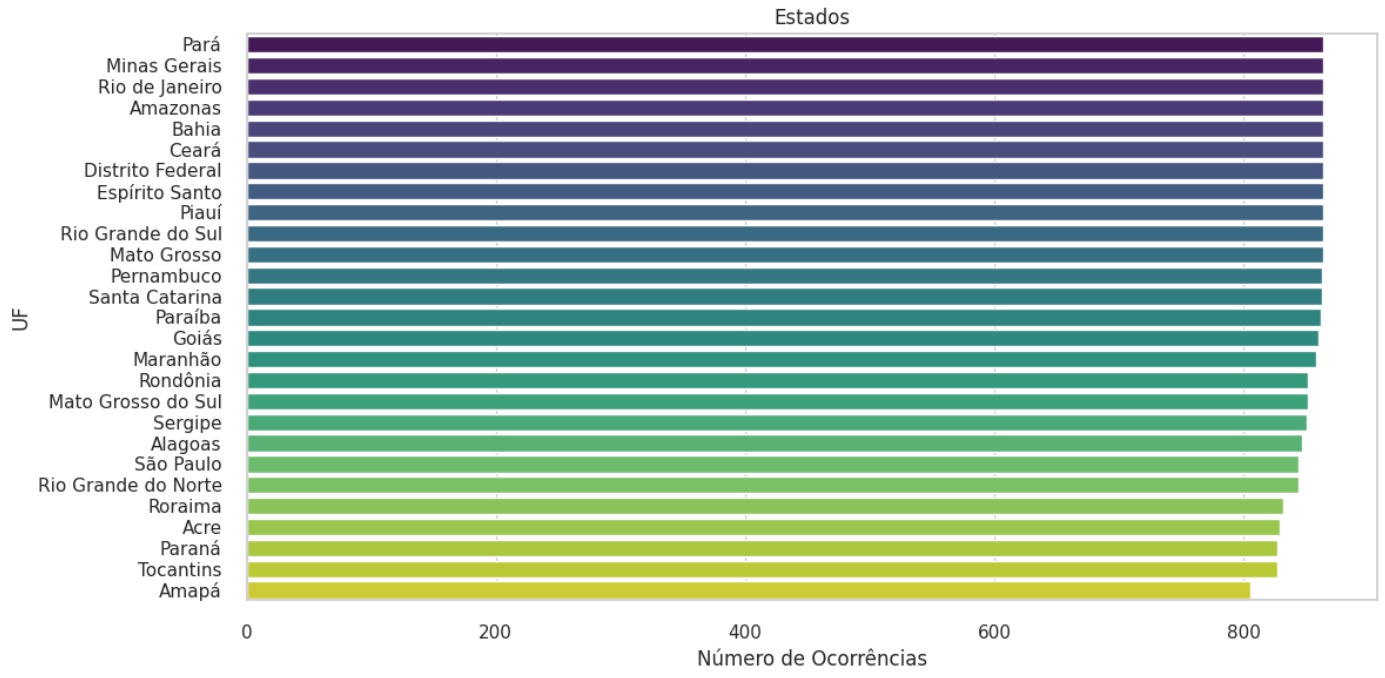
*ANEXO C*

****

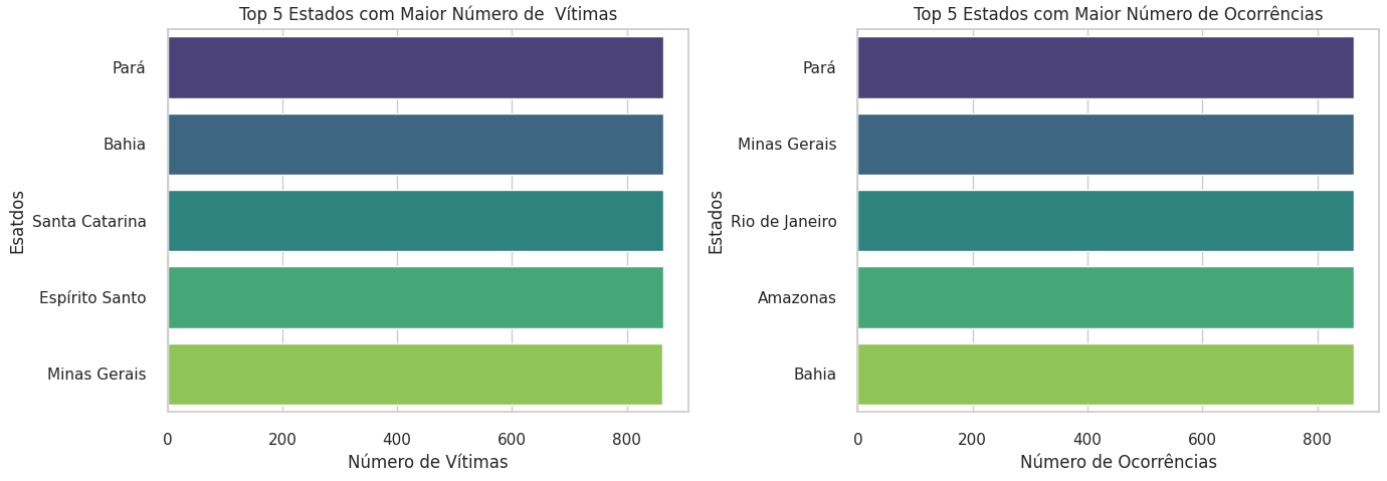
*ANEXO D*

****

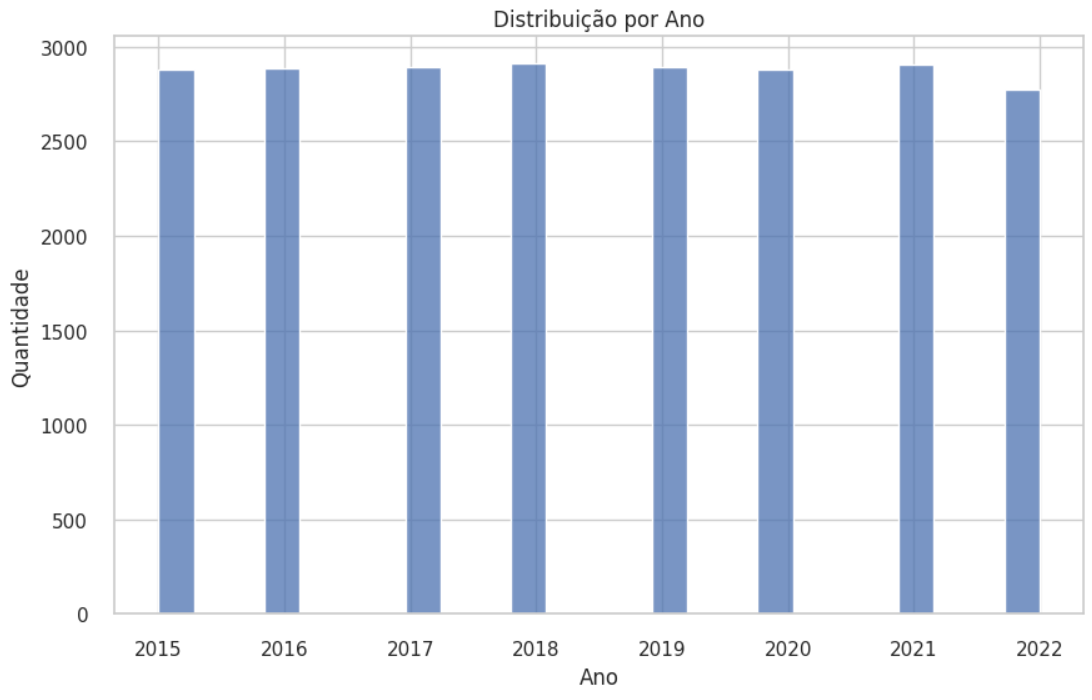
*ANEXO E*

****

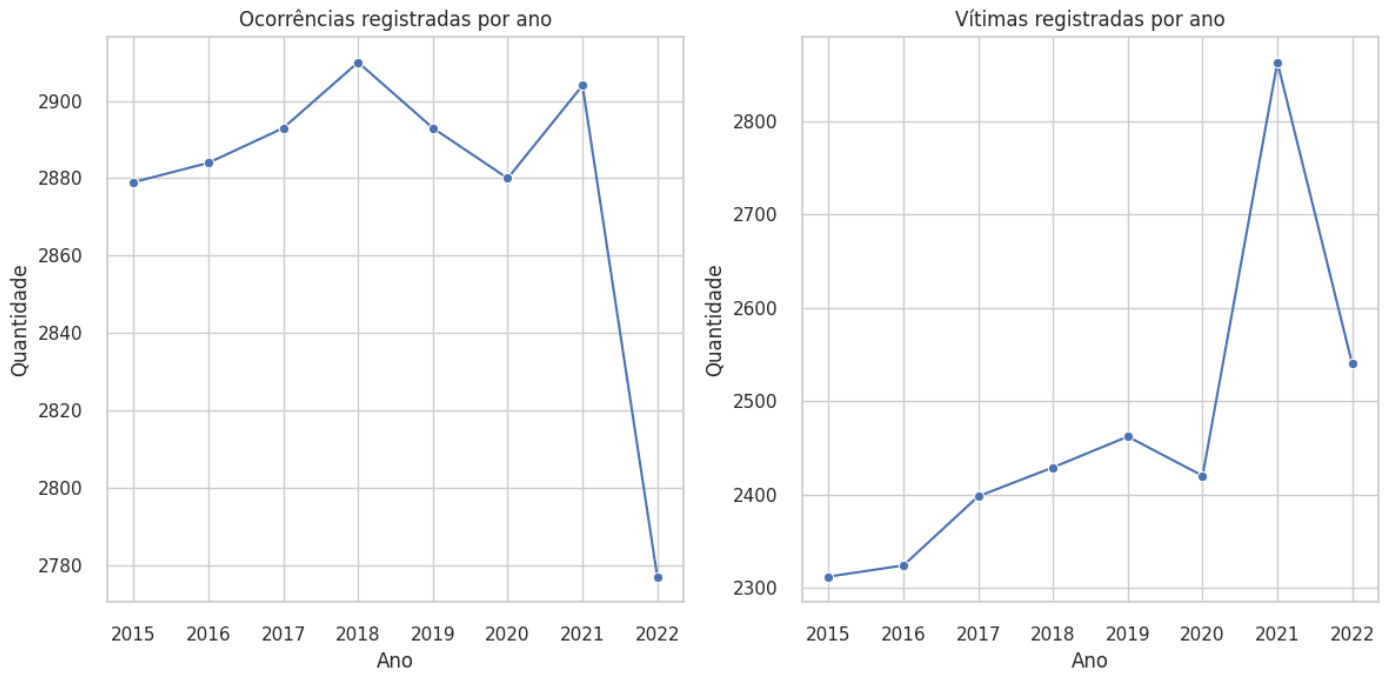
*ANEXO F*

****

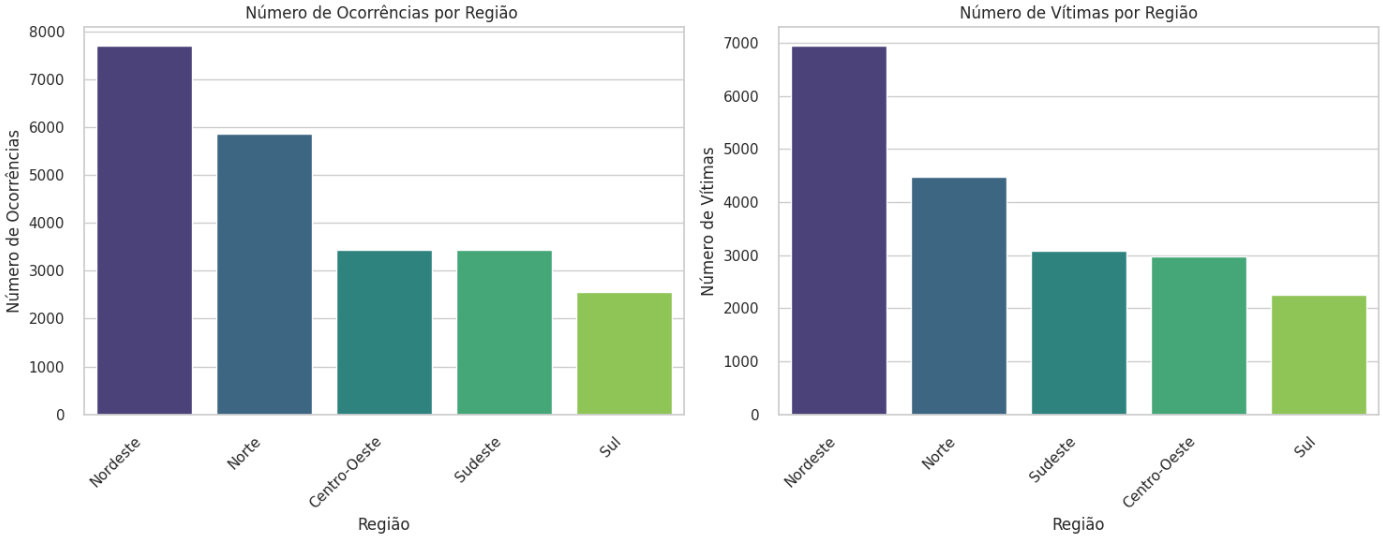
*ANEXO G*

****

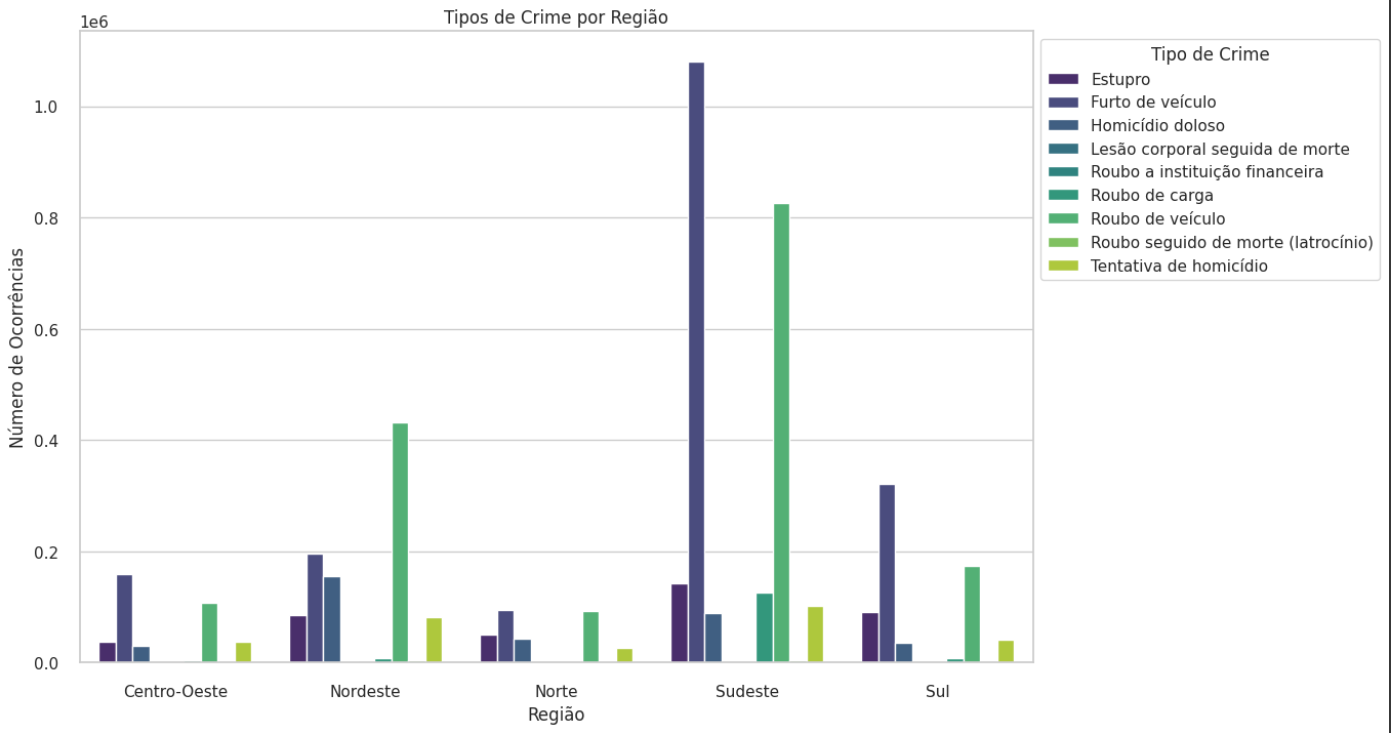
*ANEXO H*

****

*ANEXO I*

****

*ANEXO J*



# **4. Discussão:**

**4.1. Interpretação dos Resultados:**

**Homicídio Doloso e suas Implicações:** A constatação de que a maioria das ocorrências com um maior número de vítimas está associada a homicídios dolosos destaca a gravidade e a letalidade desses eventos. Isso evidencia a urgência de estratégias específicas de prevenção e intervenção para lidar com esse tipo de crime.

**Furto de Veículo e Predominância Regional:** O destaque do furto de veículos, especialmente na região sudeste, sugere a necessidade de medidas direcionadas à segurança veicular nessa área. A concentração desse tipo de crime pode indicar falhas específicas nos sistemas de prevenção ou patrulhamento.

**Perfil de Vítimas e Desafios na Identificação de Sexo:** A predominância de vítimas masculinas é um dado relevante, porém, a ausência de identificação de sexo em muitos casos indica a necessidade de uma abordagem mais inclusiva. Considerando o cenário de feminicídio, homofobia e crimes contra minorias, a inclusão de dados mais abrangentes é vital para uma análise mais precisa e atenta às disparidades sociais.

**Distribuição Regional das Ocorrências e Vítimas:** As regiões do Pará, Bahia e Minas Gerais destacam-se como áreas com maior incidência de crimes e número de vítimas. Esse padrão regional pode ser resultado de fatores socioeconômicos, demográficos e culturais, indicando a necessidade de abordagens específicas em cada localidade.

**Variação Temporal e Reflexos nas Estratégias de Segurança:** A identificação do ano de 2018 como o período com o maior número de ocorrências, seguido por uma diminuição em 2022, sugere mudanças temporais nos padrões criminais. Essa variação pode estar relacionada a políticas de segurança implementadas ou a fatores socioeconômicos e demográficos que influenciam a dinâmica criminal.

**Disparidades Regionais e Tamanho do Território:** A concentração de ocorrências e vítimas no Nordeste em comparação ao Sul pode ser parcialmente atribuída ao tamanho relativo das regiões. No entanto, isso não exclui a necessidade de análise mais detalhada das causas subjacentes a essas disparidades e a implementação de estratégias adaptadas a cada contexto regional.

**4.2. Limitações:**

**Viés nos Dados:** A presença de viés nos dados pode comprometer a representatividade da análise. Se houver subnotificação de crimes ou se certos grupos são mais propensos a relatar ocorrências, isso pode distorcer as conclusões. A falta de dados demográficos detalhados também pode introduzir viés, especialmente em relação a grupos minoritários.

**Falta de Informações Detalhadas:** A ausência de informações detalhadas sobre algumas ocorrências, como a não identificação do sexo das vítimas, limita a compreensão abrangente do cenário criminal. A falta de dados específicos pode dificultar a formulação de estratégias eficazes, especialmente em casos de crimes motivados por questões de gênero, orientação sexual ou outras características.

**Limitações nos Métodos de Coleta de Dados:** Dependendo dos métodos de coleta de dados utilizados, pode haver subjetividade ou falta de uniformidade na categorização de crimes. Variações nos critérios de registro podem afetar a consistência dos dados, dificultando comparações precisas entre regiões ou períodos.

**Sazonalidade e Mudanças Socioeconômicas:** A análise temporal pode não capturar completamente a complexidade das mudanças socioeconômicas ao longo do tempo. Fatores externos, como recessões econômicas ou mudanças nas políticas de segurança, podem influenciar os padrões criminais e não serem totalmente refletidos na análise.

**Enfoque em Crimes Reportados:** A análise se baseia em dados de crimes reportados, o que pode não refletir totalmente a realidade do crime não reportado. A subnotificação é uma preocupação, especialmente em casos de crimes que as vítimas podem hesitar em relatar, como violência doméstica ou crimes contra grupos minoritários.

**Limitações nos Modelos de Machine Learning:** A eficácia dos modelos de machine learning depende da qualidade e representatividade dos dados de treinamento. Se os dados não abrangem adequadamente a diversidade dos cenários criminais, os modelos podem ser menos precisos ou generalizáveis.

**Complexidade da Causa e Efeito:** A análise pode identificar correlações, mas estabelecer causalidade requer considerações mais aprofundadas. Fatores subjacentes, como desigualdades sociais, políticas públicas e outros elementos contextuais, podem não ser totalmente capturados pela análise e influenciar significativamente os padrões criminais.

# **5. Conclusões e Recomendações**

**5.1. Conclusões:**

O projeto de Ciência de Dados oferece uma visão abrangente do cenário criminal no Brasil, revelando padrões que requerem atenção prioritária. Homicídios dolosos e furtos de veículos surgem como os crimes mais frequentes, destacando a necessidade de estratégias preventivas específicas para lidar com sua gravidade e impacto na sociedade.

A análise revela desafios significativos na identificação de vítimas, com predominância de vítimas masculinas. A ausência de identificação de sexo em muitos casos sugere a importância de uma abordagem mais inclusiva para compreender e prevenir crimes relacionados a gênero, reconhecendo a complexidade de fenômenos como feminicídio e homofobia.

As disparidades regionais são evidentes, com Pará, Bahia e Minas Gerais apresentando maior incidência de ocorrências e vítimas. A região Sudeste destaca-se pelo predomínio de furtos de veículos, indicando a necessidade de estratégias específicas de segurança veicular nessa área.

Além dos insights valiosos, o projeto identifica limitações, como viés nos dados e falta de informações detalhadas. Essas limitações apontam para a necessidade contínua de aprimoramento das metodologias de coleta de dados e na consideração de fatores sociais mais amplos.

Em síntese, as conclusões fornecem uma base sólida para o desenvolvimento de políticas públicas informadas, estratégias de segurança preventiva e alocação eficiente de recursos. Reconhecendo as limitações, este projeto serve como um ponto de partida essencial para compreender e abordar os desafios específicos do cenário criminal brasileiro, promovendo, assim, um ambiente mais seguro e resiliente para a sociedade.

**5.2. Recomendações:**

* Desenvolver estratégias preventivas específicas para homicídios dolosos e furtos de veículos, considerando a gravidade e a frequência desses crimes. Isso pode incluir patrulhamento intensificado, medidas de segurança veicular e intervenções comunitárias.
* Aprimorar a coleta de dados, assegurando a identificação precisa de sexo e outros dados sensíveis. Isso permitirá uma compreensão mais completa dos padrões criminais, especialmente em casos relacionados a gênero, contribuindo para a formulação de políticas mais inclusivas.
* Implementar análises temporais mais dinâmicas para entender melhor as variações ao longo do tempo. Essa abordagem permitirá antecipar possíveis tendências e ajustar estratégias de segurança de acordo com as mudanças nos padrões criminais.
* Desenvolver estratégias de segurança pública adaptadas a cada região, considerando as disparidades identificadas. Isso pode incluir políticas específicas para o Pará, Bahia e Minas Gerais, levando em conta fatores socioeconômicos, culturais e demográficos únicos.
* Investir em programas educativos e preventivos em parceria com comunidades locais. Essa abordagem pode incluir iniciativas que promovam a conscientização sobre segurança, reduzindo a vulnerabilidade a crimes e fortalecendo a relação entre a comunidade e as forças de segurança.
* Implementar tecnologias avançadas, como sistemas de rastreamento veicular e vigilância inteligente, para combater furtos de veículos. Essas ferramentas podem fortalecer as estratégias de prevenção e aumentar a eficácia na recuperação de veículos roubados.
* Investir em melhorias contínuas nos métodos de coleta de dados, buscando reduzir viés e aumentar a precisão. Isso garantirá uma base de dados mais confiável para análises futuras e para orientar decisões em políticas de segurança pública.
* Promover a colaboração entre diferentes órgãos governamentais, incluindo Ministério da Justiça, segurança pública estadual e municipalidades. A integração de esforços e recursos pode potencializar a eficácia das estratégias implementadas.

# **6. Implementação e Próximos Passos:**

**6.1. Implementação das Recomendações:**

**Desenvolvimento de Estratégias Preventivas:** Estabelecer equipes multidisciplinares de especialistas em segurança pública para desenvolver estratégias específicas para homicídios dolosos e furtos de veículos. Implementar programas de patrulhamento intensificado em áreas críticas e campanhas de conscientização sobre segurança veicular.

**Melhoria na Coleta de Dados Sensíveis:** Atualizar os formulários de registro de ocorrências para garantir a identificação precisa de dados sensíveis, incluindo sexo. Oferecer treinamentos aos profissionais responsáveis pela coleta de dados para sensibilizá-los sobre a importância da inclusão correta dessas informações.

**Análise Temporal Dinâmica:** Implementar ferramentas de análise de séries temporais que permitam a identificação de padrões emergentes e variações significativas. Estabelecer protocolos para revisão periódica dessas análises, garantindo a adaptação rápida das estratégias de segurança às mudanças identificadas.

**Estratégias Regionais Adaptadas:** Criar comitês regionais de segurança pública para analisar as conclusões específicas de cada região e desenvolver estratégias personalizadas. Envolver autoridades locais, comunidades e organizações não governamentais no processo de planejamento e implementação.

**Programas Educativos e Preventivos:** Implementar campanhas educativas em escolas, bairros e comunidades, abordando temas como prevenção de crimes, segurança pessoal e formas de denúncia. Colaborar com organizações locais para envolver a comunidade no desenvolvimento e execução desses programas.

**Integração de Tecnologia na Prevenção de Furtos:** Estabelecer parcerias com empresas de tecnologia para implementar sistemas de rastreamento veicular em áreas propensas a furtos. Desenvolver políticas de incentivo para a instalação de dispositivos de segurança veicular avançados, como alarmes e bloqueadores.

**Fortalecimento da Coleta de Dados:** Investir em treinamentos regulares para os profissionais responsáveis pela coleta de dados, garantindo a uniformidade nos procedimentos. Implementar sistemas de auditoria para monitorar a qualidade e consistência dos dados, realizando correções quando necessário.

**Colaboração entre Órgãos Governamentais:** Estabelecer reuniões regulares entre representantes do Ministério da Justiça, órgãos estaduais de segurança pública e autoridades municipais. Promover a integração de informações, recursos e estratégias para uma abordagem mais coordenada e eficiente.

**6.2. Trabalhos Futuros:**

Essas áreas representam pontos-chave para o aprimoramento contínuo do entendimento do cenário criminal no Brasil. Abordar esses tópicos em análises e desenvolvimentos futuros contribuirá para estratégias mais informadas, abrangentes e adaptáveis às dinâmicas sociais e criminais em evolução.

* Uma análise mais aprofundada dos crimes relacionados a gênero, incluindo feminicídios e crimes motivados por orientação sexual, é crucial para compreender as nuances dessas ocorrências. Isso pode envolver a coleta mais detalhada de dados e a colaboração com organizações especializadas para garantir a inclusão adequada de informações sensíveis.
* A compreensão das causas subjacentes aos padrões criminais requer uma análise mais detalhada dos fatores socioeconômicos. Investigações sobre desigualdades, acesso à educação, oportunidades de emprego e outros determinantes sociais são essenciais para desenvolver estratégias holísticas que abordem as raízes do crime.
* O desenvolvimento contínuo dos modelos de machine learning utilizados pode aumentar a precisão das previsões e análises. Isso inclui a incorporação de novos dados, ajustes nos parâmetros dos algoritmos e a exploração de métodos mais avançados, como redes neurais, para lidar com a complexidade dos padrões criminais.
* A inclusão de dados mais abrangentes sobre características sociais e culturais das comunidades pode fornecer insights adicionais. Aspectos como dinâmicas familiares, influências culturais e políticas locais desempenham papéis fundamentais no entendimento do crime e devem ser considerados em futuras análises.
* Realizar avaliações sistemáticas do impacto das intervenções de segurança pública ao longo do tempo. Isso permitirá uma análise mais precisa da eficácia das estratégias implementadas, identificando aquelas que demonstram resultados positivos e aquelas que requerem ajustes ou revisão.
* Investigar a subnotificação de crimes e a percepção da comunidade sobre a segurança. Entender por que alguns crimes não são reportados e como a comunidade percebe as estratégias de segurança pública pode informar iniciativas de conscientização e confiança entre as autoridades e a população.
* Desenvolver análises específicas para diferentes faixas etárias, considerando que padrões criminais podem variar entre jovens e idosos. Isso proporcionará insights específicos para a criação de programas preventivos direcionados a grupos etários específicos, contribuindo para uma abordagem mais eficaz.

# **7. Referências:**

Ministério da Justiça e Segurança Pública do Brasil. (2023). "Ocorrências Criminais - Sinesp". Disponível em: <https://dados.gov.br/dados/conjuntos-dados/sistema-nacional-de-estatisticas-de-seguranca-publica> . Acesso em 15 de junho de 2024.

Dados.gov.br. "Sistema Nacional de Estatísticas de Segurança Pública". Disponível em: <https://dados.mj.gov.br/dataset/210b9ae2-21fc-4986-89c6-2006eb4db247/resource/fca1cb9b-1bfb-4090-bdca-5171ed3b9fa4/download/dicionario-de-dados-convertido.pdf> Acesso em 15 de junho de 2024.

# 