

Mackenzie

Curso Tecnológico em Ciências de Dados

Projeto Aplicado I

Bruno Ferreira - RA: 10443074

João Pineda - RA: 10433696

Lucas Feo Mazzei - RA: 10433452

Rian Fernandes - RA: 10433740

**Análise de Dados de Homicídios Dolosos no Estado de São Paulo: Uma
Abordagem Baseada em dados da Secretaria de Segurança Pública.**

São Paulo

2024

Índice

Introdução.....	3
Cronograma	3
1. Organização.....	4
2. Apresentação dos Dados.	4
3. Proposta Analítica e Resultados pretendidos.	5
3. Link para o Projeto no GitHub.....	7
4. Análise Exploratória dos Dados.	7
4.1 Informações do Dataset.....	7
4.2 Valores Nulos	8
4.3 Principais Estatísticas	8
4.4 Idade das Vítimas	9
4.5 Ano da Ocorrência.....	10
4.6 Hora da Ocorrência.....	11
4.7 Município da Ocorrência	12
4.7 Tipo de Local	13
4.8 Sexo das Vítimas	13
4.9 Cor da Pele.....	14
4.10 Profissão das Vítimas	14
Roteiro do Storytelling.....	15
1. Setup	15
2. Conflito.....	16
3. Ponto de virada.....	16
4. Resolução.....	17

Introdução

Este trabalho foi realizado por um grupo de estudantes com o objetivo de analisar dados de homicídios dolosos no estado de São Paulo. Os dados utilizados estão disponíveis no portal de transparência da Secretaria de Segurança Pública do Estado de São Paulo (SSP-SP) e abrangem o período de 2017 a 2022.

O projeto busca entender os padrões e tendências dos homicídios dolosos no estado, utilizando uma base de dados extensa que inclui informações sobre localização, data e hora dos crimes, características das vítimas (como idade, gênero, raça e profissão) e detalhes específicos dos incidentes. A análise desses dados permite identificar grupos mais vulneráveis e compreender melhor as dinâmicas da criminalidade.

Ao realizar essa análise, o grupo espera contribuir para uma compreensão mais aprofundada dos fatores que influenciam os homicídios dolosos em São Paulo, fornecendo insights que possam ser utilizados para o desenvolvimento de políticas de segurança pública mais eficazes e estratégias de prevenção.

Cronograma

Atividades	Meses											
	Setembro				Outubro				Novembro			
Seleção do Dataset e Proposta Inicial de análise.	X	X	X									
Tratamento das bases.		X	X	X	X							
Análises Descritivas.				X	X	X	X					
Gravação do Vídeo.							X	X	X	X		

1. Organização

Para desenvolver o projeto foi escolhida a Secretaria da Segurança Pública do Estado de São Paulo (SSP-SP) que é um órgão do governo estadual responsável pela coordenação e administração das políticas de segurança pública. A SSP-SP tem como principais funções a prevenção, controle e combate à criminalidade, além de garantir a segurança das pessoas, propriedades e comunidades.

A SSP-SP também coordena programas de segurança comunitária e trabalha em conjunto com outras instituições para promover a segurança e o bem-estar da população.

Essa secretaria conta com uma base de dados extensa que contém informações sobre diversos setores onde a organização atua.

2. Apresentação dos Dados.

Os dados escolhidos estão disponíveis no portal de transparência da SSP¹ e refere-se aos registros de homicídios dolosos do estado de São Paulo entre os períodos de 2017 a 2022.

A Secretaria de Segurança Pública de São Paulo divulga bases de registros de crimes contra a vida e as divide entre homicídio doloso, feminicídio, latrocínio, morte suspeita e lesão corporal seguida de morte, para o projeto em questão, foi escolhida a base 'Homicídio Doloso' diante da robustez dos dados – tendo em vista que, nesta base, as informações à respeito das vítimas e do fato ocorrido são mais completas, além de contar com mais registros que as demais dada a natureza generalista do crime.

¹ <https://www.ssp.sp.gov.br/estatistica/spvida>

3. Proposta Analítica e Resultados pretendidos.

A partir da análise dos dados disponibilizados, buscamos identificar padrões, aferir tendências ao longo do tempo, relacionar informações de gênero, profissão e faixa etária para compreender os grupos mais vulneráveis à condição de vítima desses crimes.

Abaixo, buscamos definir cada uma das colunas do conjunto de dados a fim de compreender, na integralidade, o conjunto de dados.

Nome	Definição	Tipo
Departamento da Circunscrição	Departamento responsável pelo registro do homicídio.	String
Seccional da Circunscrição	Divisão da seccional do departamento responsável.	String
Município da Circunscrição	Município Responsável pelo B.O.	String
Departamento da Circunscrição	Subdivisão do município geralmente, bairro.	String
Número de Vítima de Homicídio Doloso	Quantidade de vitimados pelo ato registrado.	Int.
Identificação da Delegacia	ID único da delegacia responsável pela apuração.	Int.
Mês da Estatística	Mês de registro da ocorrência.	Int.
Ano da Estatística	Ano de registro da ocorrência.	Int.
Data e Hora do Registro do Boletim de Ocorrência	Data completa do registro.	Date
Número do Boletim de Ocorrência	ID único do boletim de ocorrência.	String
Ano do Boletim de Ocorrência	Ano de registro do Boletim de Ocorrência	Int.
Município de Elaboração	Município Responsável pelo B.O.	String

Departamento de Elaboração	Nome da delegacia responsável pela elaboração	String
Seccional de Elaboração	Divisão da seccional do departamento responsável pelo B.O.	String
Departamento de Elaboração	Departamento responsável pelo registro do homicídio.	String
Data do Fato	Dia do ocorrido presente no B.O.	Date
Hora do Fato	Hora relatada no B.O.	Timestamp
Descrição do Tipo de Local	Local relatado da ocorrência.	String
Logradouro	Endereço do ocorrido.	String
Número do Logradouro	Número do local caso trate-se de uma edificação.	Int.
Latitude	Coordenadas do local onde o fato ocorreu.	Float
Longitude	Coordenadas do local onde o fato ocorreu.	Float
Tipo de Pessoa	Qualificação do indivíduo focaremos em vítimas.	String
Sexo da Pessoa	Gênero do indivíduo.	String
Idade da Pessoa	Anos completos de idade.	Int.
Data de Nascimento da Pessoa	Data de nascimento completa do indivíduo.	Date
Cor da Pele	Raça do indivíduo.	String
Profissão	Ocupação profissional do indivíduo.	String
Natureza Apurada	Homicídio doloso.	String

3. Link para o Projeto no GitHub.

<https://github.com/jpopineda10433696/Projeto-Aplicado-I.git>

4. Análise Exploratória dos Dados.

A partir do script em Python desenvolvido no Jupyter Notebook e disponibilizado no GitHub mencionado, foi possível realizar diversas análises significativas sobre as ocorrências de homicídios dolosos registrados em São Paulo entre 2017 e 2022.

4.1 Informações do Dataset

Informações sobre nomes das colunas do dataset, tipos de variáveis, presença de valores nulos e número de linhas e colunas.

```
df.info()

<class 'pandas.core.frame.DataFrame'>
Index: 18436 entries, 0 to 3843
Data columns (total 33 columns):
 #   Column                                  Non-Null Count  Dtype
---  -
 0   DEPARTAMENTO_CIRCUNSCRICAO             18436 non-null  object
 1   SECCIONAL_CIRCUNSCRICAO               18436 non-null  object
 2   MUNICIPIO_CIRCUNSCRICAO               18436 non-null  object
 3   DP_CIRCUNSCRICAO                     18436 non-null  object
 4   HD                                     17528 non-null  float64
 5   Nº DE VÍT HD                           18436 non-null  int64
 6   ID_DELEGACIA                          18436 non-null  object
 7   MÊS ESTATISTICA                       18436 non-null  int64
 8   ANO ESTATISTICA                       18436 non-null  int64
 9   DATAHORA_REGISTRO_BO                 18436 non-null  datetime64[ns]
10  NUM_BO                                18436 non-null  object
11  ANO_BO                                18436 non-null  int64
12  MUNICIPIO_ELABORACAO                  18436 non-null  object
13  DP_ELABORACAO                        18436 non-null  object
14  SEC_ELABORACAO                       18436 non-null  object
15  DEP_ELABORACAO                       18436 non-null  object
16  DATA_FATO                           18436 non-null  datetime64[ns]
17  HORA_FATO                             16805 non-null  float64
18  DESC_TIPOLOCAL                        18430 non-null  object
19  LOGRADOURO                           18436 non-null  object
20  NUMERO_LOGRADOURO                     14151 non-null  object
21  LATITUDE                              13275 non-null  float64
22  LONGITUDE                             13276 non-null  float64
23  TIPO_PESSOA                           18436 non-null  object
24  SEXO_PESSOA                           18436 non-null  object
25  IDADE_PESSOA                          16397 non-null  float64
26  DATA_NASCIMENTO_PESSOA               16232 non-null  datetime64[ns]
27  COR_PELE                              18433 non-null  object
28  PROFISSAO                             10976 non-null  object
29  NATUREZA_APURADA                      18436 non-null  object
30  ANO                                    18436 non-null  int32
31  MÊS                                    18436 non-null  int32
32  index                                 18436 non-null  int64
dtypes: datetime64[ns](3), float64(5), int32(2), int64(5), object(18)
memory usage: 4.6+ MB
```

4.2 Valores Nulos

Número de valores nulos por coluna - será importante para verificarmos se é uma informação crítica sobre a vítima ou do crime que esteja faltando.

```
Número de valores nulos por coluna:
DEPARTAMENTO_CIRCUNSCRICAO      0
SECCIONAL_CIRCUNSCRICAO        0
MUNICIPIO_CIRCUNSCRICAO        0
DP_CIRCUNSCRICAO               0
HD                             908
Nº DE VÍT HD                   0
ID_DELEGACIA                   0
MÊS_ESTATISTICA                0
ANO_ESTATISTICA                0
DATAHORA_REGISTRO_BO           0
NUM_BO                         0
ANO_BO                         0
MUNICIPIO_ELABORACAO           0
DP_ELABORACAO                  0
SEC_ELABORACAO                 0
DEP_ELABORACAO                 0
DATA_FATO                      0
HORA_FATO                      0
DESC_TIPOLOCAL                 6
LOGRADOURO                     0
NUMERO_LOGRADOURO              4285
LATITUDE                       5161
LONGITUDE                      5160
TIPO_PESSOA                    0
SEXO_PESSOA                    0
IDADE_PESSOA                   2029
DATA_NASCIMENTO_PESSOA        2205
COR_PELE                       3
PROFISSAO                     7462
NATUREZA_APURADA               0
dtype: int64
```

4.3 Principais Estatísticas

Nesta etapa, exploramos as principais medidas estatísticas de cada coluna da base de dados, utilizando todas as funções fornecidas pelo método `describe()` do `pandas`. Esse método permite uma visão abrangente e detalhada dos dados, fornecendo as seguintes estatísticas:

- Contagem (count): quantidade de valores não nulos em cada coluna, o que permite entender a completude dos dados.
- Média (mean): valor médio dos dados, oferecendo uma indicação da centralidade dos valores para variáveis numéricas.

- Desvio padrão (std): uma medida da dispersão dos dados, indicando o grau de variação ou dispersão em relação à média.
- Valor mínimo (min): o menor valor encontrado em cada coluna, fornecendo um limite inferior dos dados.
- Quartis (25%, 50%, 75%): também conhecidos como o primeiro, segundo (mediana) e terceiro quartis, que ajudam a entender a distribuição dos valores em cada coluna.
- Valor máximo (max): o maior valor presente na coluna, indicando o limite superior dos dados.

	HD	Nº DE VÍT	MÊS	ESTATISTICA	ANO	ESTATISTICA	DATAHORA_REGISTRO_BO	ANO_BO	DATA_FATO	LATITUDE	LONGITUDE	IDADE_PESSOA	DATA_NASCIMENTO_PESSOA	ANO	MÊS	index
count	17528.0	18436.0	18436.000000	18436.000000	18436.000000	18436.000000	18436.000000	18436.000000	18436.000000	13275.000000	13276.000000	16397.000000	16232.000000	18436.000000	18436.000000	18436.000000
mean	1.0	1.0	6.494413	2019.420048	2019.420048	2019-12-02 00:16:58.679772672	2019.421241	2019-12-01 09:27:22.612280320	-23.063920	-47.043924	35.304873	1984-02-13 19:41:13.336619008	2019.419234	6.494305	1544.634411	
min	1.0	1.0	1.000000	2017.000000	2017.000000	2017-01-01 00:00:00	2017.000000	2017-01-01 00:00:00	-25.013276	-52.946693	0.000000	1900-01-04 00:00:00	2017.000000	1.000000	1.000000	
25%	1.0	1.0	3.000000	2018.000000	2018.000000	2018-05-06 00:00:00	2018.000000	2018-05-05 00:00:00	-23.613403	-47.317517	25.000000	1976-04-05 12:00:00	2018.000000	3.000000	769.000000	
50%	1.0	1.0	6.000000	2019.000000	2019.000000	2019-12-03 00:00:00	2019.000000	2019-12-02 00:00:00	-23.453237	-46.754710	34.000000	1985-09-11 00:00:00	2019.000000	6.000000	1537.000000	
75%	1.0	1.0	10.000000	2021.000000	2021.000000	2021-05-31 17:02:30	2021.000000	2021-05-31 00:00:00	-22.799619	-46.476299	43.000000	1994-05-25 06:00:00	2021.000000	10.000000	2305.250000	
max	1.0	1.0	12.000000	2022.000000	2022.000000	2023-01-01 00:00:00	2023.000000	2022-12-31 00:00:00	0.000000	0.000000	99.000000	2022-05-07 00:00:00	2022.000000	12.000000	3504.000000	
std	0.0	0.0	3.535203	1.737525	NaN	1.739483	NaN	0.885320	1.291725	13.552341	NaN	1.737675	3.536538	900.311048		

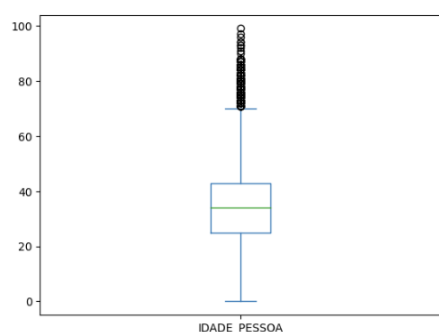
4.4 Idade das Vítimas

Medidas estatísticas sobre a idade das vítimas presentes no relatório. A partir desta visualização, conseguimos identificar que a idade média das vidas é de 35 anos – valor muito próximo da mediana (34), o que nos leva a determinar que estejamos num cenário de distribuição normal (ou de sino) dos valores.

Ademais, é importante salientar a presença de outliers ao observar dados de vítimas com idades mais avançadas (aproximadamente acima dos 70 anos)

IDADE_PESSOA	
count	16397.000000
mean	35.304873
std	13.552341
min	0.000000
25%	25.000000
50%	34.000000
75%	43.000000
max	99.000000

Construção de boxplot para verificar a distribuição de idades da coluna 'IDADE_PESSOA'.

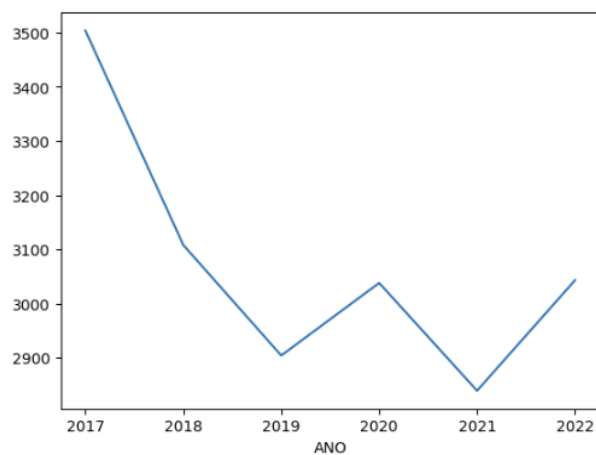


4.5 Ano da Ocorrência

Contagem de valores por ano de ocorrência do crime. É possível observar uma queda no número de homicídios no estado de SP desde o ano de 2017, com um pico no ano de 2020 e tendência de aumento em 2022.

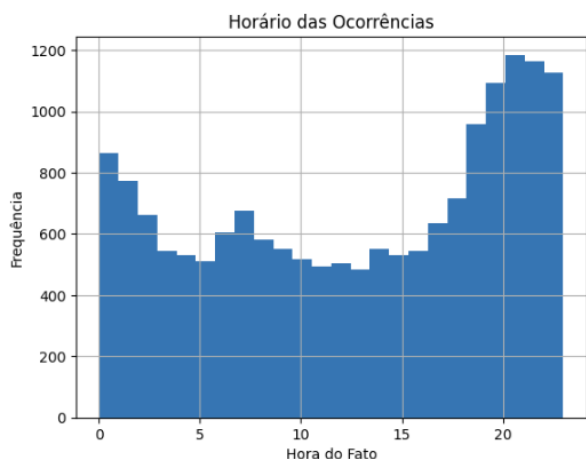
	count
ANO	
2017	3504
2018	3108
2022	3043
2020	3038
2019	2904
2021	2839

Abaixo, plotamos um gráfico de linhas para observar a evolução temporal do número de crimes por ano de ocorrência:



4.6 Hora da Ocorrência

Contagem de valores por hora da ocorrência. É observável que a maioria dos homicídios que ocorreram no período 2017-2022 foram no período da noite, com destaque dos horários entre 20h00 e 23h00.



4.7 Município da Ocorrência

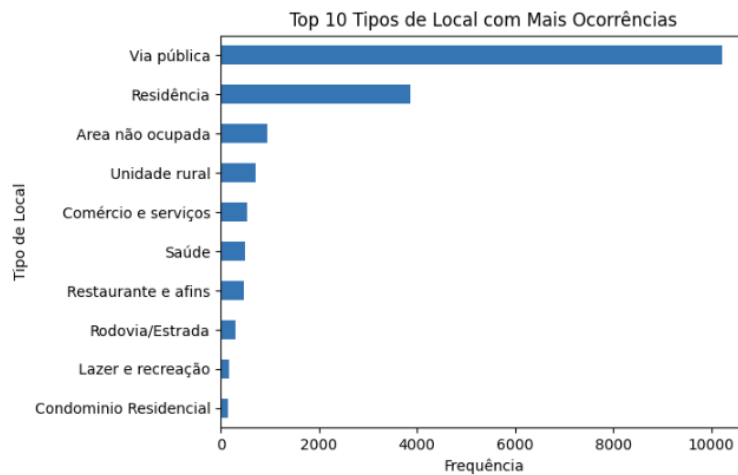
Ocorrência dos crimes por município. O número mais expressivo de homicídios ocorridos no período selecionado foi na cidade de São Paulo, representando cerca de 4000 ocorrências no total.

É importante salientar que os 5 primeiros municípios com maior número de crimes ocorridos estão dentre as 7 cidades mais populosas do estado de São Paulo, o que influencia no número absoluto de homicídios apresentados acima.



4.7 Tipo de Local

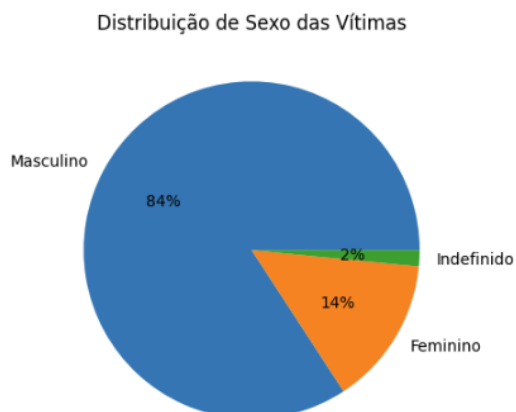
Top 10 locais com maior número de homicídios. É possível observar que a maior parte dos homicídios ocorridos entre 2017-2022 foram em via pública, mais do que o dobro das ocorrências em locais residenciais.



4.8 Sexo das Vítimas

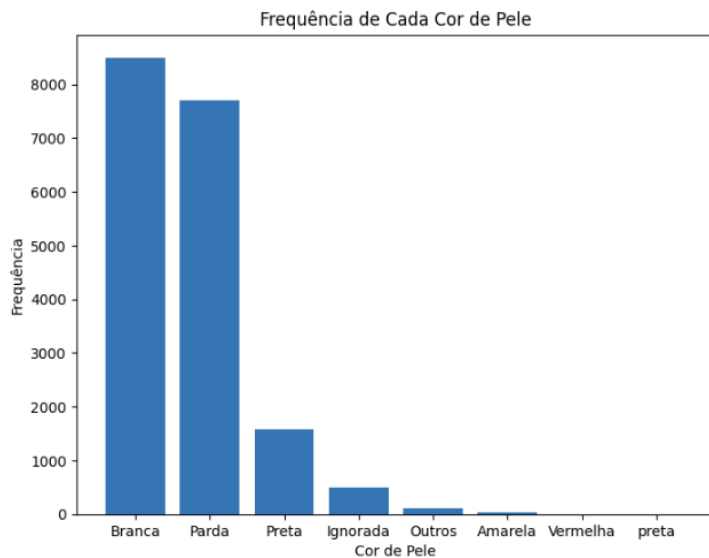
Distribuição percentual de ocorrência por sexo da vítima.

Dentre os homicídios observados, 84% têm como vítimas pessoas do sexo masculino, seguido por 14% do sexo feminino.



4.9 Cor da Pele

Dentre todos os dados observados, mais de 8.500 homicídios ocorridos foram de pessoas de pele branca, seguidos de cerca de 7.700 casos de crimes contra pessoas pardas.



4.10 Profissão das Vítimas

Retirando os casos em que as informações de profissão não foram disponibilizadas, as três maiores ocorrências de homicídios no estado de São Paulo estão distribuídas entre vítimas desempregadas (1224), estudantes (1032) e autônomos(as) (497).



Roteiro do Storytelling

1. Setup

Nos últimos anos, a cidade de São Paulo tem um desafio considerável: o alto número de homicídios dolosos, que prejudicam a qualidade de vida dos cidadãos e limitam o bem-estar e o sentimento de segurança de muitas pessoas. Além de vítimas fatais, esses atos criminosos custam à sociedade um alto custo social e econômico ao sobrecarregar o sistema de justiça e segurança pública e desestabilizar comunidades inteiras.

Portanto, é crucial identificar as causas subjacentes dessa tendência e os fatores que colaboram para ela. Com os dados de ocorrências desde 2001 fornecidos pela Secretaria de Segurança Pública de São Paulo, este trabalho pretende estudar e analisar os padrões de violência, impondo um perfil das vítimas e locais de instalação para ajudar a diagnosticar a complexa dinâmica por trás da violência.

Essas informações são essenciais para desenvolver iniciativas e políticas de segurança melhores e mais abrangentes, que possam reduzir a violência e garantir um ambiente de segurança e respeito à dignidade humana para todos.

2. Conflito

As taxas alarmantes de homicídios dolosos em São Paulo representam um sério desafio à qualidade de vida dos cidadãos. Esses crimes violentos não apenas geram um sinal de alerta para as autoridades, mas também causam um sentimento de insegurança generalizada na população, o que evidencia a urgência de compreender suas causas e dinâmicas.

Para enfrentar esses conflitos, é fundamental identificar grupos vulneráveis e os fatores que contribuem para a ocorrência desses crimes. Além do sofrimento humano, os homicídios resultam em custos sociais e econômicos significativos, incluindo a sobrecarga dos sistemas de justiça e policial e a desestabilização do meio social, afetando negativamente a qualidade de vida da sociedade.

Com os dados fornecidos pela Secretaria de Segurança Pública de São Paulo (SSP-SP), é possível traçar perfis detalhados das vítimas, analisar as características dos crimes e mapear os locais de maior incidência de homicídio dolosos, trazendo informações cruciais para identificar fatores de risco e dinâmicas que influenciam a violência, criando ferramentas que permitam direcionar estratégias de prevenção e combate ao crime.

Compreender as causas dessa problemática e suas implicações é essencial para desenvolver ações efetivas que reduzam a violência e promovam um ambiente mais seguro e que respeite a dignidade humana.

3. Ponto de virada

Durante a análise exploratória dos dados de homicídios em São Paulo fornecidos pela Secretaria de Segurança Pública de São Paulo (SSP-SP), descobrimos padrões e tendências que ajudam a entender a magnitude e as nuances do problema.

A taxa de homicídios variou significativamente ao longo dos anos, com picos em determinados períodos, indicando possíveis relações com eventos sociais ou mudanças em políticas públicas. É revelado uma tendência de alta em 2022. Nota-se um pico nos casos de homicídio entre 20 hrs e 23 hrs.

Áreas específicas da cidade apresentaram taxas mais altas de homicídios, sugerindo uma concentração do problema nas cidades mais populosas do estado, tendo como principal cidade, a cidade de São Paulo. Os crimes são cometidos em sua maioria em vias públicas.

A maioria das vítimas de homicídios foram adultos brancos e pardos do sexo masculino com média de 35 anos e alguns outliers em torno dos 70 anos.

A maioria das vítimas constam como desempregadas, o que pode sugerir conflitos causados pelo aumento da violência causado pelo desemprego.

Os dados foram coletados a partir dos registros da Secretaria de Segurança Pública de São Paulo (SSP-SP), garantindo uma base sólida para a análise. Cada registro incluía informações detalhadas sobre a vítima, o local, o método e o contexto do crime, permitindo uma visão abrangente e multifacetada do problema.

4. Resolução

A análise dos homicídios dolosos no estado de São Paulo entre 2017 e 2022 revela importantes insights sobre as vítimas, os locais, e os contextos dos crimes. O estudo exploratório das métricas demonstrou que, apesar de uma tendência de redução nos homicídios ao longo dos anos, o ano de 2020 apresentou um aumento significativo, possivelmente influenciado por fatores socioeconômicos ou conjunturais específicos - levando em consideração que este fora o ano de maior impacto da pandemia do coronavírus que acarretou numa perda econômica severa.

Contudo, mesmo com uma leve redução nos anos subsequentes, 2022 mostra sinais de crescimento nos registros, sugerindo uma possível necessidade de reforço em estratégias de segurança e políticas preventivas.

Em relação ao perfil das vítimas, os dados mostram que os crimes atingem principalmente a população adulta, tendo em vista a distribuição normal da faixa etária com pico próximo aos 35 anos e de profissão 'autônomo' ou desempregados - o que evidencia a necessidade de políticas inclusivas como forma de prevenção à situações de risco.

A maior parte dos homicídios ocorreram à noite, especialmente entre 20h e 23h, em vias públicas e isto nos mostra uma necessidade de reforçar a segurança preventiva, como policiamento em áreas públicas e iluminação adequada para reduzir o número de crimes desta espécie.

Desta forma, essa análise demonstra que, ao conhecer padrões de ocorrência de crimes e características das vítimas, é possível direcionar esforços de prevenção e segurança pública com mais precisão e eficácia.