

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

ANDERSON NOGUEIRA SILVA
BRENO MOURA DE ABREU
GIOVANA VIEGAS BARROS
HEITOR DERDER TREVISOL

ATIVIDADE 02 - ANÁLISE EXPLORATÓRIA

INTRODUÇÃO A CIÊNCIA DE DADOS

CURITIBA

2023

ANDERSON NOGUEIRA SILVA
BRENO MOURA DE ABREU
GIOVANA VIEGAS BARROS
HEITOR DERDER TREVISOL

ATIVIDADE 02 - ANÁLISE EXPLORATÓRIA

Relatório Parcial da disciplina Introdução a Ciência de Dados, apresentado ao professor que a ministra na Universidade Tecnológica Federal do Paraná como requisito parcial para obtenção de aprovação na disciplina.

Orientador: Prof. Dr. Luiz Celso Gomes Junior

CURITIBA

2023

LISTA DE FIGURAS

FIGURA 1	– Histograma de Ocorrências por Ano.	8
FIGURA 2	– Variação percentual do número de ocorrências anual.	8
FIGURA 3	– Histograma de Ocorrências por Mês.	9
FIGURA 4	– Variação percentual do número de ocorrências mensal.	10
FIGURA 5	– Histograma de ocorrências por dia da semana.	10
FIGURA 6	– Variação percentual do número de ocorrências por dia da semana.	11
FIGURA 7	– Histograma de ocorrências por hora do dia.	12
FIGURA 8	– Variação percentual do número de ocorrências por hora do dia.	12
FIGURA 9	– Histograma do número de ocorrências para os 10 bairros com mais ocorrências.	13
FIGURA 10	– Histograma igual ao da figura 9, porém sem o bairro Centro.	14
FIGURA 11	– Mapa de calor de ocorrências atendidas pela GM em Curitiba.	15
FIGURA 12	– O mesmo mapa da figura 11, porém sem o bairro Centro.	16
FIGURA 13	– Quantidade total para cada tipo de crime	18
FIGURA 14	– Quantidade de crimes por bairro	19
FIGURA 15	– Quantidade das 20 maiores ocorrências agrupadas por crime e bairro	20
FIGURA 16	– Quantidade de crimes relacionados a dano por dia da semana	21
FIGURA 17	– Quantidade de crimes relacionados a dano por mês	21
FIGURA 18	– Quantidade de crimes relacionados a dano por ano	22
FIGURA 19	– Quantidade de crimes relacionados a substâncias ilícitas por dia da semana	22
FIGURA 20	– Quantidade de crimes relacionados a substâncias ilícitas por mês	23
FIGURA 21	– Quantidade de crimes relacionados a substâncias ilícitas por ano	23
FIGURA 22	– Quantidade de crimes relacionados a roubo por dia da semana	24
FIGURA 23	– Quantidade de crimes relacionados a roubo por mês	24
FIGURA 24	– Quantidade de crimes relacionados a roubo por ano	25
FIGURA 25	– Quantidade de crimes relacionados a furto por dia da semana	25
FIGURA 26	– Quantidade de crimes relacionados a furto por mês	26
FIGURA 27	– Quantidade de crimes relacionados a furto por ano	26
FIGURA 28	– Quantidade de ocorrências por dia da semana para o primeiro grupo	27
FIGURA 29	– Quantidade de crimes por tipo para o primeiro grupo	27
FIGURA 30	– Quantidade de ocorrências por dia da semana para o segundo grupo	27
FIGURA 31	– Quantidade de crimes por tipo para o segundo grupo	27
FIGURA 32	– Quantidade de ocorrências por dia da semana para o terceiro grupo	27
FIGURA 33	– Quantidade de crimes por tipo para o terceiro grupo	28
FIGURA 34	– Óbitos por Mês de ocorrência	29
FIGURA 35	– Óbitos por Agressão por Mês de ocorrência	29
FIGURA 36	– Óbitos por Intervenção Legal por Mês de ocorrência	30
FIGURA 37	– Óbitos por Ano de ocorrência	31
FIGURA 38	– Óbitos por Agressão por Ano de ocorrência	31
FIGURA 39	– Óbitos por Intervenção Legal por Ano de ocorrência	32
FIGURA 40	– Óbitos Indeterminados por Ano de ocorrência	32

FIGURA 41 – Resumo da quantidade de homicídios em Curitiba	33
FIGURA 42 – Homicídios em Curitiba por ano	33
FIGURA 43 – Resumo da quantidade de mortes violentas por causa indeterminada em Curitiba	34
FIGURA 44 – Mortes violentas por causa indeterminada em Curitiba por ano	34

SUMÁRIO

1	INTRODUÇÃO	3
1.1	TEMA	3
1.2	EQUIPE	3
2	DESENVOLVIMENTO	4
2.1	OBTENÇÃO E PROCESSAMENTO DE DADOS	4
2.1.1	SiGesGuarda	4
2.1.2	DataSUS	5
2.1.3	IPEA Atlas da Violência	6
2.2	COBERTURA E DISTRIBUIÇÃO DOS DADOS	7
2.2.1	SiGesGuarda	7
2.2.1.1	Visualizações Gerais	7
2.2.2	DataSUS	28
2.2.3	IPEA (Homicídios)	33
2.2.4	IPEA (Mortes Violentas por Causa Indeterminada)	34
2.3	IDENTIFICAÇÃO DE PADRÕES INICIAIS	35
2.3.1	SiGesGuarda	35
2.3.2	DataSus	35
2.3.3	IPEA	35
2.4	PERGUNTAS DE PESQUISA E EXPLORAÇÕES INICIAIS	35
2.5	CONCLUSÃO E PRÓXIMOS PASSOS	37
	REFERÊNCIAS	38

1 INTRODUÇÃO

1.1 TEMA

Estudo de tendências na criminalidade na cidade de Curitiba com análises regionais e temporais. O objetivo principal da equipe é entender como está situada a distribuição de crimes e ocorrências policiais, levando em consideração a região, o tipo do crime e dados temporais registrados. Além disso, também temos o objetivo de verificarmos as tendências de criminalidade em Curitiba, permitindo estimar o estado futuro das ocorrências.

1.2 EQUIPE

Detetives de Dados

Integrantes:

Anderson Nogueira Silva, 2126516

Breno Moura de Abreu, 1561286

Giovana Viegas Barros, 2090066

Heitor Derder Trevisol, 1611810

2 DESENVOLVIMENTO

Nesta etapa, foram obtidos dados de diversas fontes diferentes, estas sendo:

- SisGeGuarda (TRÂNSITO, 2022)
- DataSUS (SAÚDE, 2020)
- IPEA Atlas da Violência (Homicídios) (IPEA, 2019)
- IPEA Atlas da Violência (Mortes Violentas por Causa Indeterminada) (IPEA, 2019)
- Casos de COVID-19 em Curitiba (Prefeitura Municipal de Curitiba, 2023)

Como estes dados foram adquiridos de fontes diferentes possuem, portanto, padrões diferentes. Desta forma, foram utilizados métodos de limpeza e agregação de dados diferentes para cada fonte.

2.1 OBTENÇÃO E PROCESSAMENTO DE DADOS

A forma de obtenção dos dados e a descrição de como os dados foram tratado estão descritos nessa seção para cada uma das bases de dados.

2.1.1 SIGESGUARDA

Esta fonte de dados é o foco principal de análise da equipe. Sendo assim, análises mais abrangentes foram realizadas sobre este conjunto de dados. Os dados podem ser encontrados na página da prefeitura municipal da cidade de Curitiba, especificamente no Portal de Dados Abertos (Prefeitura Municipal de Curitiba, acesso em 16 abril 2023).

De modo sucinto, as informações contidas neste conjunto de dados, são relativas às ocorrências atendidas pela Guarda Municipal de Curitiba. Possui ocorrências desde o ano de

2009 até o ano de 2022. Ainda, na página de *download* dessas informações, é indicado que a base de dados possui uma frequência de atualizações mensal.

As informações obtidas no Portal de Dados Abertos foram carregadas em formato .csv (Comma Separated Values), isto é, valores separados por vírgulas, e todo o seu processamento foi realizado em Jupyter Notebook (KLUYVER et al., 2016) com a linguagem de programação Python (ROSSUM; DRAKE, 2009). Entre as bibliotecas de processamento e visualizações de dados utilizadas, estão: pandas (Pandas Development Team, 2021), geopandas (GeoPandas Development Team, 2021) e matplotlib (Matplotlib Development Team, 2021).

Para o processamento dessa base de dados, inicialmente, optou-se por realizar uma verificação na qualidade dos valores presentes em cada coluna. Além disso, foi realizada uma validação de quais colunas continham as informações de maior relevância para a análise pretendida. Todas as informações presentes nesta base de dados podem ser encontradas no dicionário de dados específico deste conjunto de dados (Prefeitura Municipal de Curitiba, 2015).

Com as colunas mais relevantes definidas, iniciou-se um tratamento individual das informações contidas em cada uma dessas colunas, para validar a qualidade das informações presentes no *dataset*. Portanto, o tratamento desses dados seguiu o detalhamento a seguir.

- Todas as colunas cujos valores eram inteiros foram convertidas para inteiros; Isto teve de ser realizado, pois ao verificar os valores únicos de certa coluna, um mesmo número inteiro aparecia duas vezes. Após a execução deste passo, os valores únicos se mostraram de forma coerente.
- Todas as colunas cujos valores eram strings que, de certa forma, representam valores categóricos, como o tipo do crime ou o bairro em que se passou a ocorrência, foram formatados para apresentar os dados apenas em letras maiúsculas. Desta forma, evitou-se que uma mesma categoria descrita em minúscula e maiúscula pudesse prejudicar as visualizações.

2.1.2 DATASUS

Os dados do DataSus utilizados foram encontradas na plataforma *TABNET*, a qual agrega dados diversos relacionados a saúde no Brasil, para está análise foram utilizados os dados referentes a óbitos por causas externas na cidade de Curitiba, Paraná. Estes dados podem ser encontrados na seção **Estatísticas Vitais**, subseção **Mortalidade - desde 1996 pela CID-10**; dentro desta seleção foi então escolhido *Óbitos por causas externas* com Abrangência geográfica *Brasil por Municípios*. Para um estudo mais abrangente foi também adquirido os

dados totais de óbitos para os mesmos períodos na cidade de Curitiba na seleção *Mortalidade Geral*.

Com os dados relevantes encontrados eles foram carregados no formato CSV e importados dentro de um arquivo pelo *Jupyter Notebook* (exp_datasus_heitor) e o processo de agregação e limpeza foi dado início.

1. Os 2 quadros de dados foram unidos utilizando a coluna referente ao período (Ano/Mês) dos óbitos como índice para a união;
2. As colunas de categoria do óbito que não são relevantes para a pesquisa foram ignoradas mantendo apenas (Total por período, Agressões, Óbitos indeterminados e Intervenção Legal);
3. As colunas foram então renomeadas para nomes padronizados, e a coluna (Ano/Mês) foi separada em 2 colunas uma referente ao ano e outra ao mês das ocorrências;
4. As linhas referentes aos totais anuais foram retiradas para uma leitura melhor da base;
5. Os campos em que nenhum óbito foi registrado no período estavam inicialmente indicados pelo carácter "-", entretanto para o estudo é preferível o algarismo "0" e portanto os dados foram alterados para serem apresentados totalmente numericamente;
6. Com esta mudança foi possível alterar a tipagem dos dados na coluna de *object* para *int64*;
7. Por fim a coluna **Mês** foi modificada para que as strings do mês tivessem o número do mês no início para uma melhor visualização ordenada.

2.1.3 IPEA ATLAS DA VIOLÊNCIA

Os dados do IPEA Atlas da Violência são conjuntos de dados relativos ao número de ocorrências criminais dos municípios do Brasil entre os anos de 2000 e 2019. No site do IPEA é possível encontrar diversas fontes de dados separadas por tipo de crime como homicídios, violência física, violência por gênero e mortes violentas por causa indeterminada. Para esta análise, foram utilizadas apenas duas fontes de dados do IPEA: homicídios e mortes violentas por causa indeterminada.

As fontes de dados são encontradas no formato CSV e possuem três colunas relevantes: Município, Ano e Valor, que indica a quantidade de ocorrências daquele tipo. A análise exploratória para os dados do IPEA podem ser visualizados nos arquivos exp_ipea_homic.ipynb e exp_ipea_indet.ipynb.

Os seguintes passos foram efetuados para a limpeza de dados:

1. Filtrar os dados apenas para o município de Curitiba;
2. Remover linhas duplicadas.

2.2 COBERTURA E DISTRIBUIÇÃO DOS DADOS

Nesta seção, estão distribuídas as visualizações dos dados geradas, durante a fase de análise exploratória, descrições e interpretações sobre o que foi obtido.

2.2.1 SIGESGUARDA

Apenas para fins de dimensionamento desta base de dados, há nela 429347 registros de ocorrências, atendidas pela Guarda Municipal de Curitiba, ao longo do período do ano de 2009 até 2022.

Desta forma, as visualizações construídas estão apresentadas nesta seção, bem como aquilo que pôde ser extraído de cada uma delas.

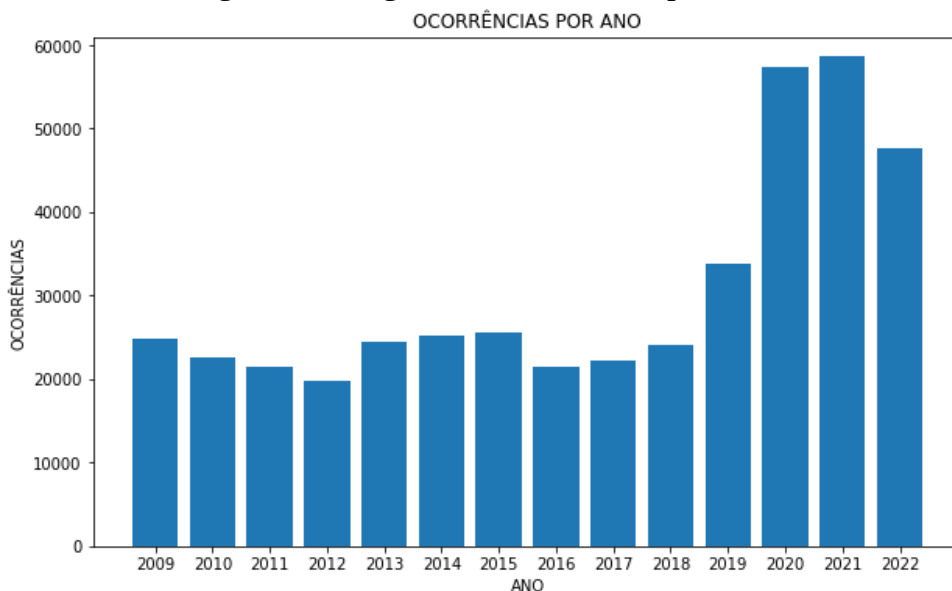
2.2.1.1 VISUALIZAÇÕES GERAIS

Nesta seção estão contidas as visualizações gerais sobre os dados contidos na base de dados.

Foram elaboradas visualizações para representar o número de ocorrências ao longo de todo o período dentro de determinado contexto. Como, por exemplo: anos, dias da semana, horas do dia, e meses do ano. A seguir podemos ver o número de ocorrências para os contextos especificados.

Na figura 1 abaixo, é possível verificar a totalidade de ocorrências atendidas pela GM Curitiba para todo o período de abrangência, isto é, desde 2009 até 2022.

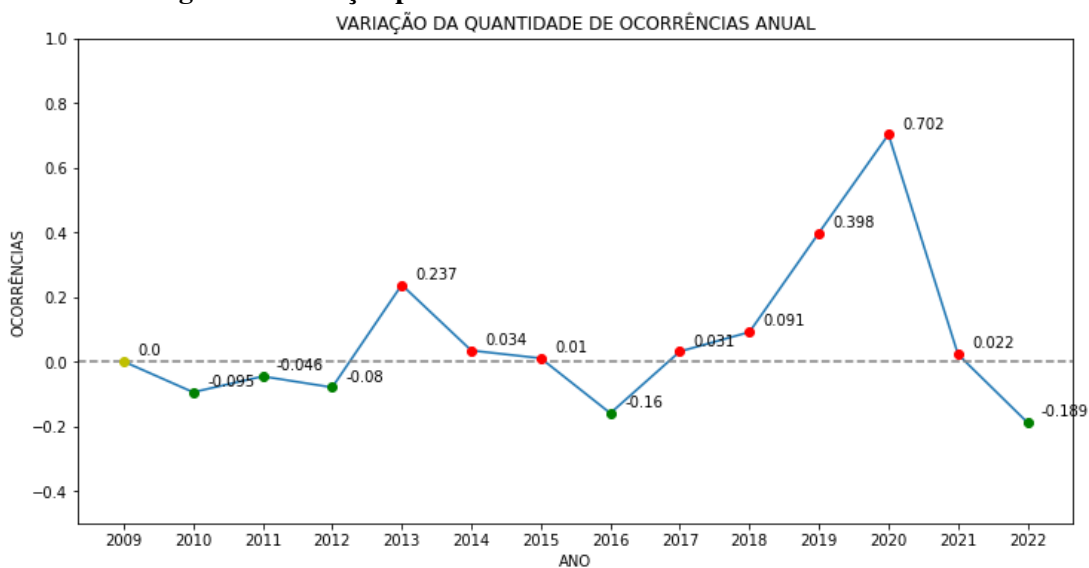
Figura 1: Histograma de Ocorrências por Ano.



Fonte: Autores.

Como complemento da figura 1, a figura 2 apresenta a variação percentual do número de ocorrências atendidas ao longo dos anos.

Figura 2: Variação percentual do número de ocorrências anual.

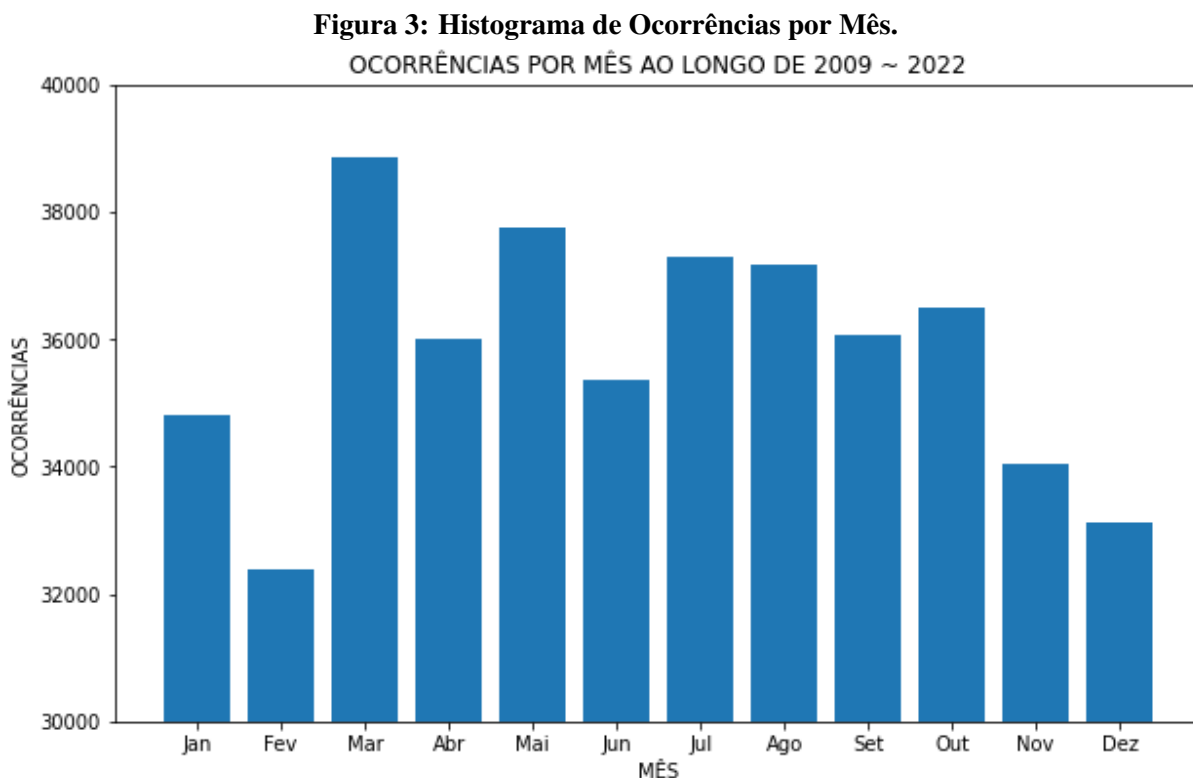


Fonte: Autores.

A partir dessas duas visualizações é possível notar que há uma tendência de aumento na criminalidade registrada pela GM de Curitiba.

Também se pode notar um aumento bastante expressivo de ocorrências do ano de 2019 para 2020, e um número bastante superior aos outros anos em 2020 e 2021. Este fato pode ser associado à pandemia por conta do COVID-19 e suas implicações sociais e econômicas.

As visualizações para o número de ocorrências por meses do ano, ao longo de 2009 até 2022, pode ser vista na figura 3.

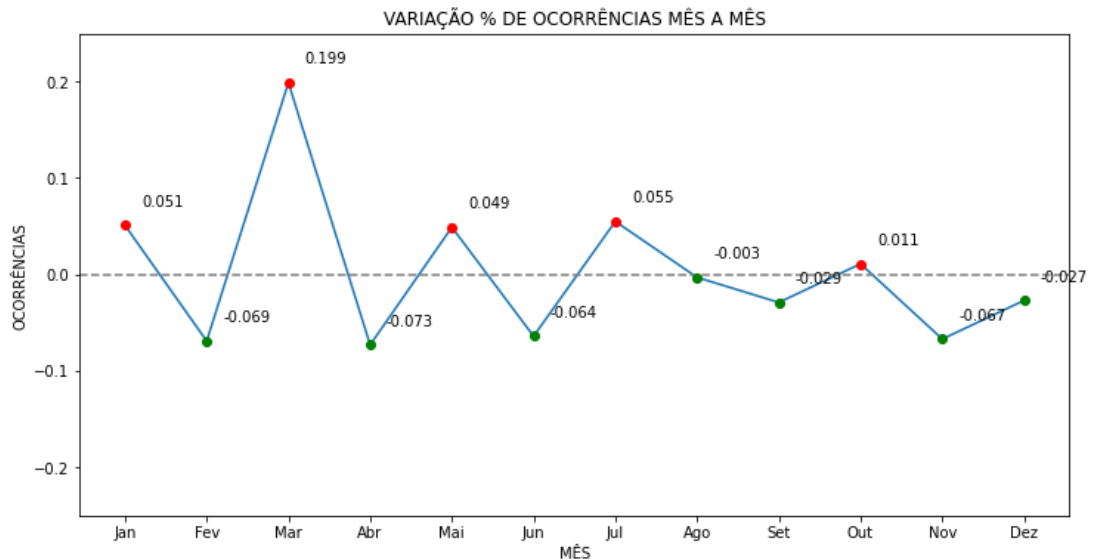


Fonte: Autores.

Com base nesta visualização, é possível averiguar que o mês de março teve o maior número de ocorrências atendidas pela GM. Já o mês de fevereiro o menor número, o que pode ser explicado por possuir apenas 28 dias, ou 29 dias em anos bissextos.

As variações percentuais do número de ocorrências ao longo dos meses do ano podem ser visualizadas na figura 4 abaixo.

Figura 4: Variação percentual do número de ocorrências mensal.

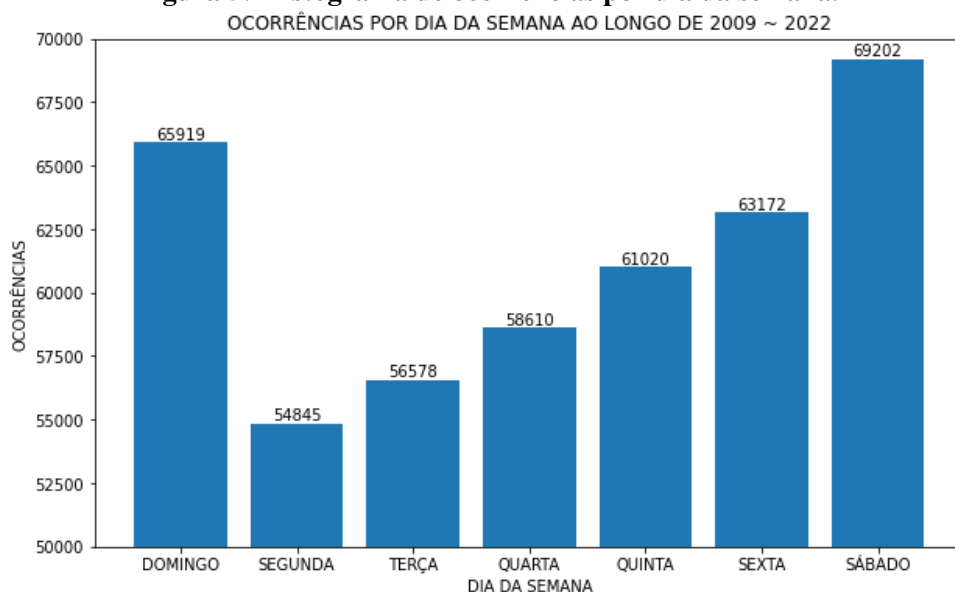


Fonte: Autores.

Com base na figura 4, é possível verificar um pico de aumento de ocorrências do mês de fevereiro para o mês de março e uma leve redução de ocorrências registradas ao longo dos meses até o final do ano, em novembro e dezembro.

A figura 5, a seguir, representa um histograma para o número de ocorrências para os dias da semana.

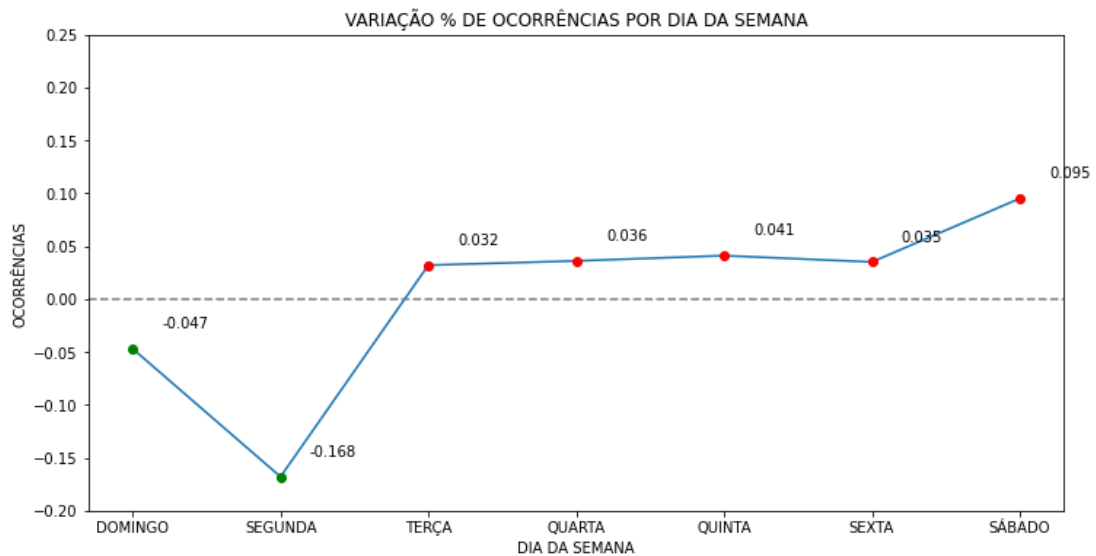
Figura 5: Histograma de ocorrências por dia da semana.



Fonte: Autores.

A figura 6 abaixo, representa as variações percentuais de um dia para outro ao longo de uma semana.

Figura 6: Variação percentual do número de ocorrências por dia da semana.

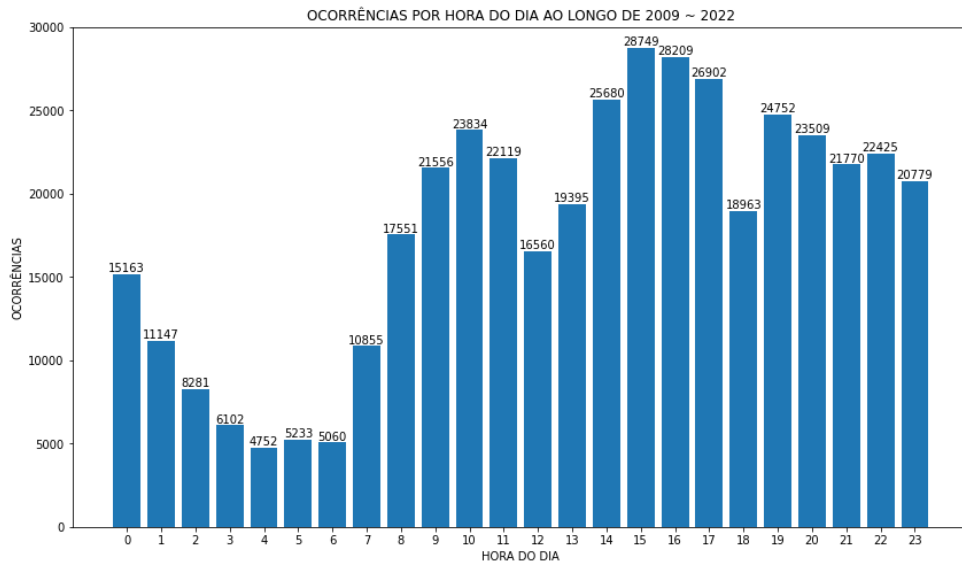


Fonte: Autores.

Os dias com maior incidência de ocorrências atendidas está nos sábados e domingos. Além disso, percebe-se um aumento ao longo da semana, que inicia na segunda-feira, e reduz drasticamente o de domingo para segunda-feira.

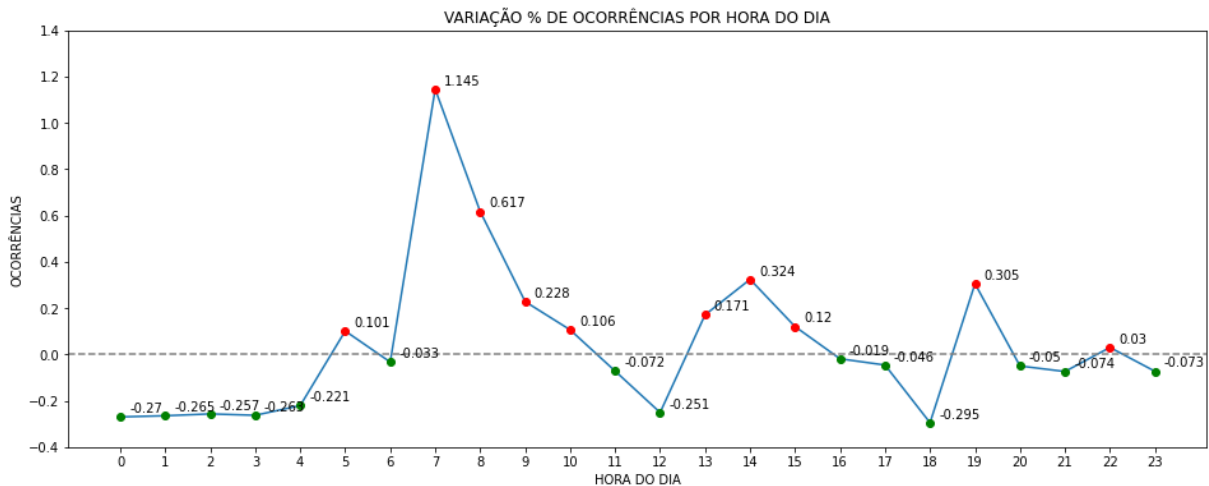
A última análise de modo análogo às anteriores, se concentra nas horas do dia em que as ocorrências foram registradas.

Portanto, a figura 7 representa o número de ocorrências que foram registradas por hora do dia ao longo do período de 2009 até 2022.

Figura 7: Histograma de ocorrências por hora do dia.

Fonte: Autores.

A figura 8, a seguir, indica a variação do número de ocorrências atendidas de hora para hora de um dia.

Figura 8: Variação percentual do número de ocorrências por hora do dia.

Fonte: Autores.

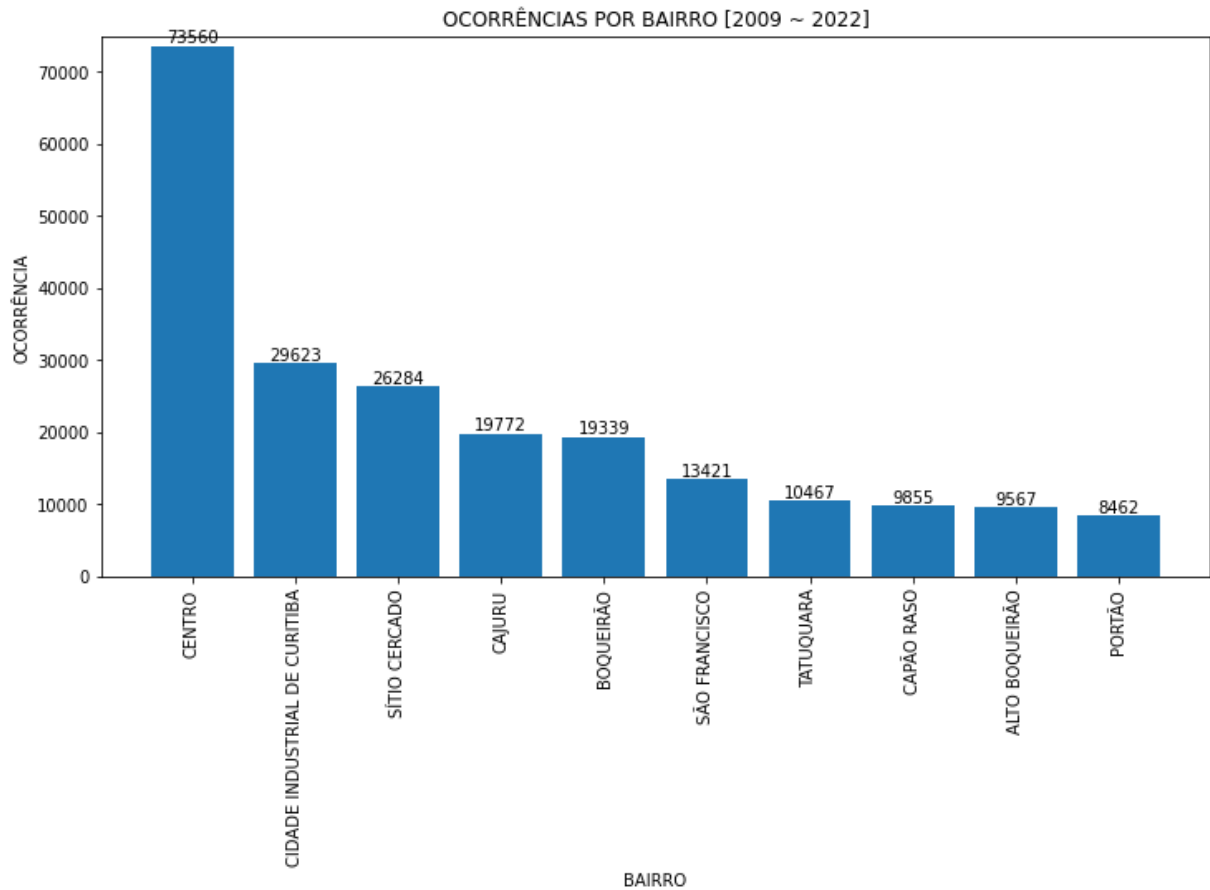
Por meio dessas visualizações é possível perceber a maior frequência de ocorrências acontece no período da tarde. Sendo que o período da noite até às 23h também demonstra um grande número de ocorrências registradas.

Outro ponto interessante são o número de ocorrências atendidas no período da madrugada, representando o momento do dia com menor frequência de ocorrências registradas.

Como uma forma de analisar geograficamente a distribuição das ocorrências, no período de 2009 até o ano de 2022, foi elaborado um histograma que indica o número de ocorrências atendidas pela GM de Curitiba para os 10 bairros com o maior número de ocorrências. Além disso, há também um mapa de calor do número de ocorrências por bairro da cidade de Curitiba. As informações geográficas da divisa de bairros de Curitiba foi obtida a partir do Instituto de Pesquisa e Planejamento urbano de Curitiba (IPPUC, 2018).

A figura 9 abaixo, indica o histograma do número de ocorrências para os 10 bairros com mais ocorrências na cidade de Curitiba.

Figura 9: Histograma do número de ocorrências para os 10 bairros com mais ocorrências.

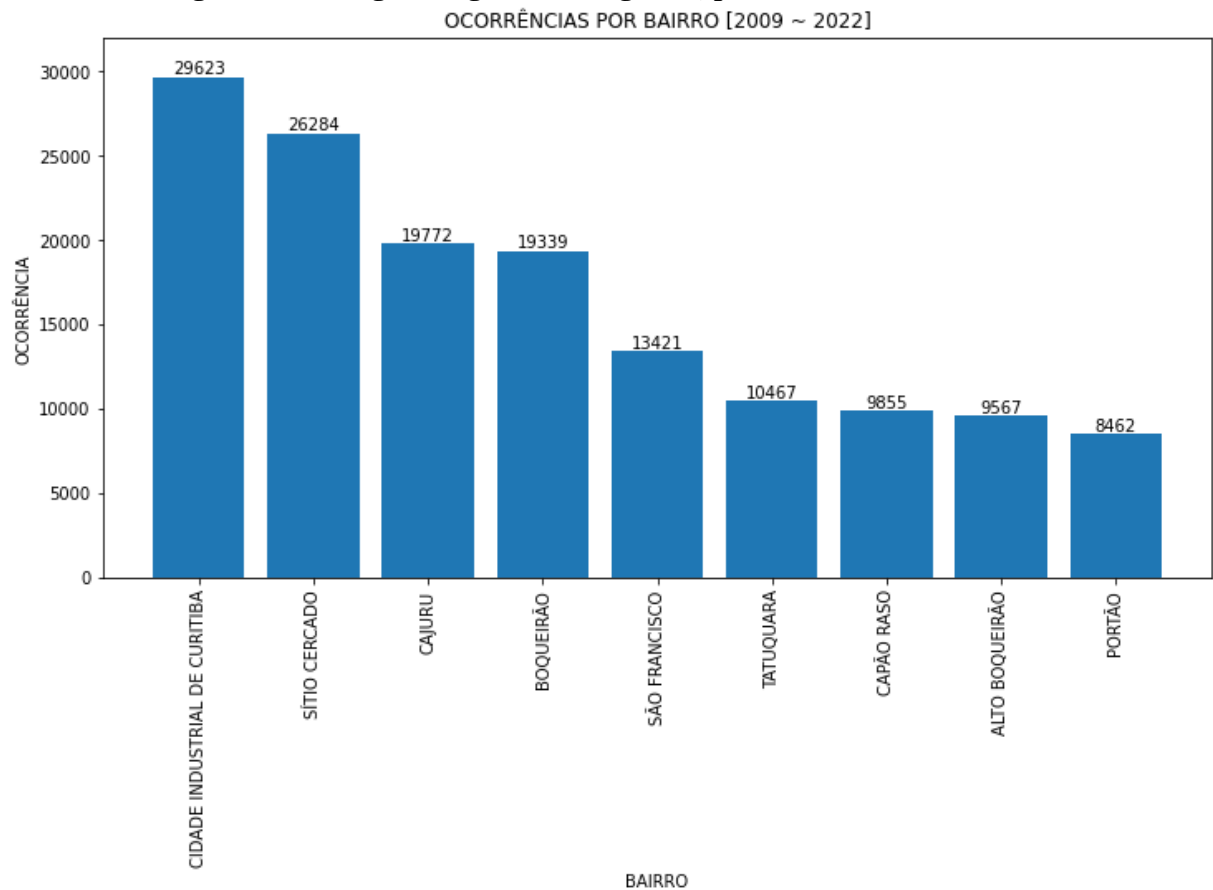


Fonte: Autores.

A visualização deste histograma demonstra que o bairro Centro é disparado o bairro com mais ocorrências.

Para facilitar a visualização, também foi gerado o mesmo gráfico que o anteriores da figura 9, porém sem considerar o bairro Centro que distorcia as escalas da visualização com seu grande número de ocorrências registradas. O histograma mencionado pode ser visualizado na figura 10.

Figura 10: Histograma igual ao da figura 9, porém sem o bairro Centro.

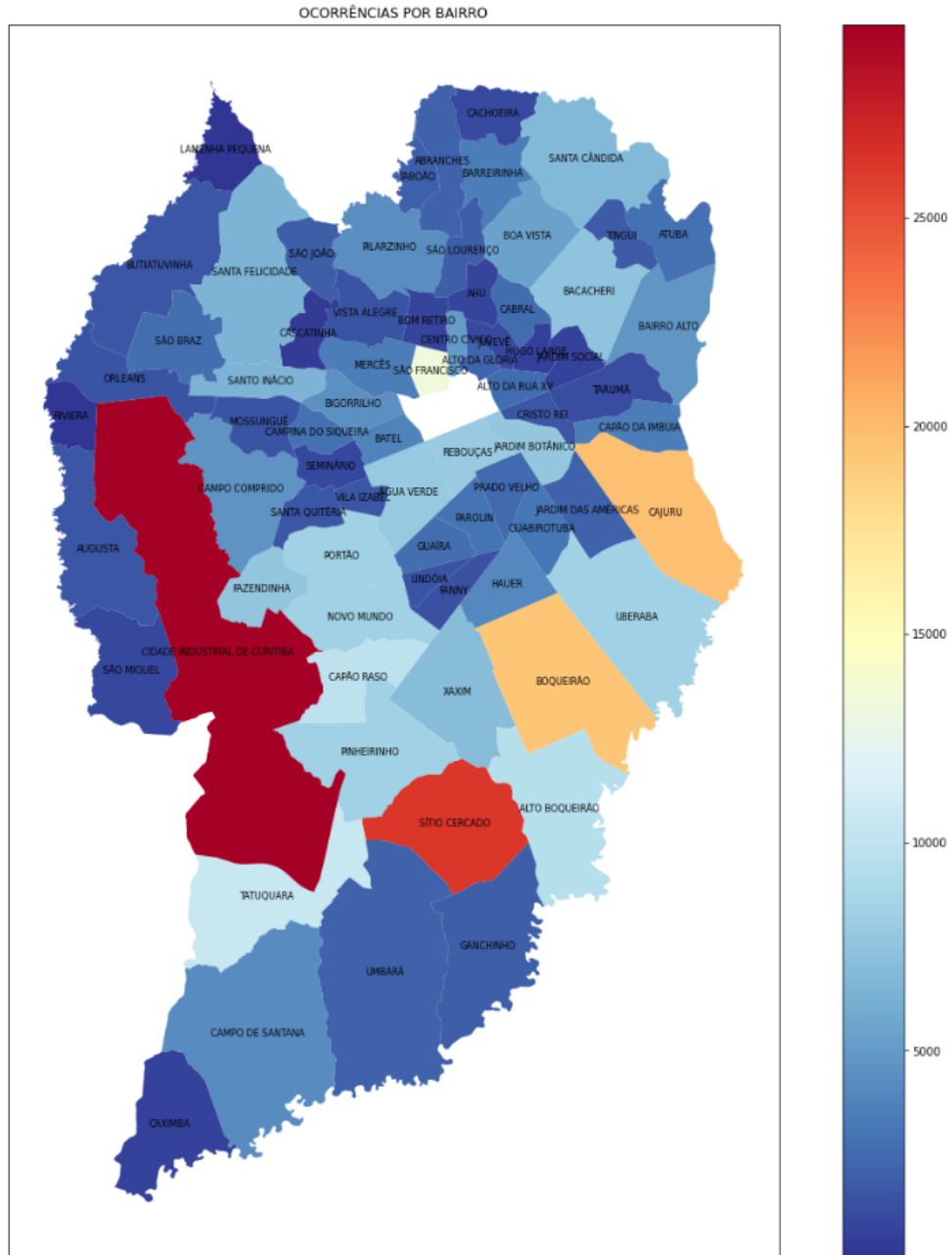


Fonte: Autores.

Com a retirada do bairro Centro do histograma, vemos que do 3º para o 4º e do 5º para o 6º bairro com mais ocorrências registradas há uma diferença significativa.

De modo semelhante com a escala do histograma da figura 9, foi gerado também o mapa de calor sem considerar o bairro Centro. Este mapa pode ser visualizado na figura 12, a seguir.

Figura 12: O mesmo mapa da figura 11, porém sem o bairro Centro.



Fonte: Autores.

Uma das análises realizadas foi feita através da seleção das linhas relacionadas com as seguintes ocorrências de natureza criminosa:

1. Agressão física/verbal
2. Crime ambiental
3. Dano
4. Estupro
5. Furto
6. Homicídio
7. Maus tratos a animais
8. Roubo
9. Sequestro e cárcere privado
10. Substância Ilícita
11. Violação de Medida Protetiva Lei Maria da Penha
12. Porte Ilegal

As linhas relativas à essas ocorrências representam 23% dos dados com um total de 99524 entradas.

As análises realizadas com os dados de ocorrências criminais foram:

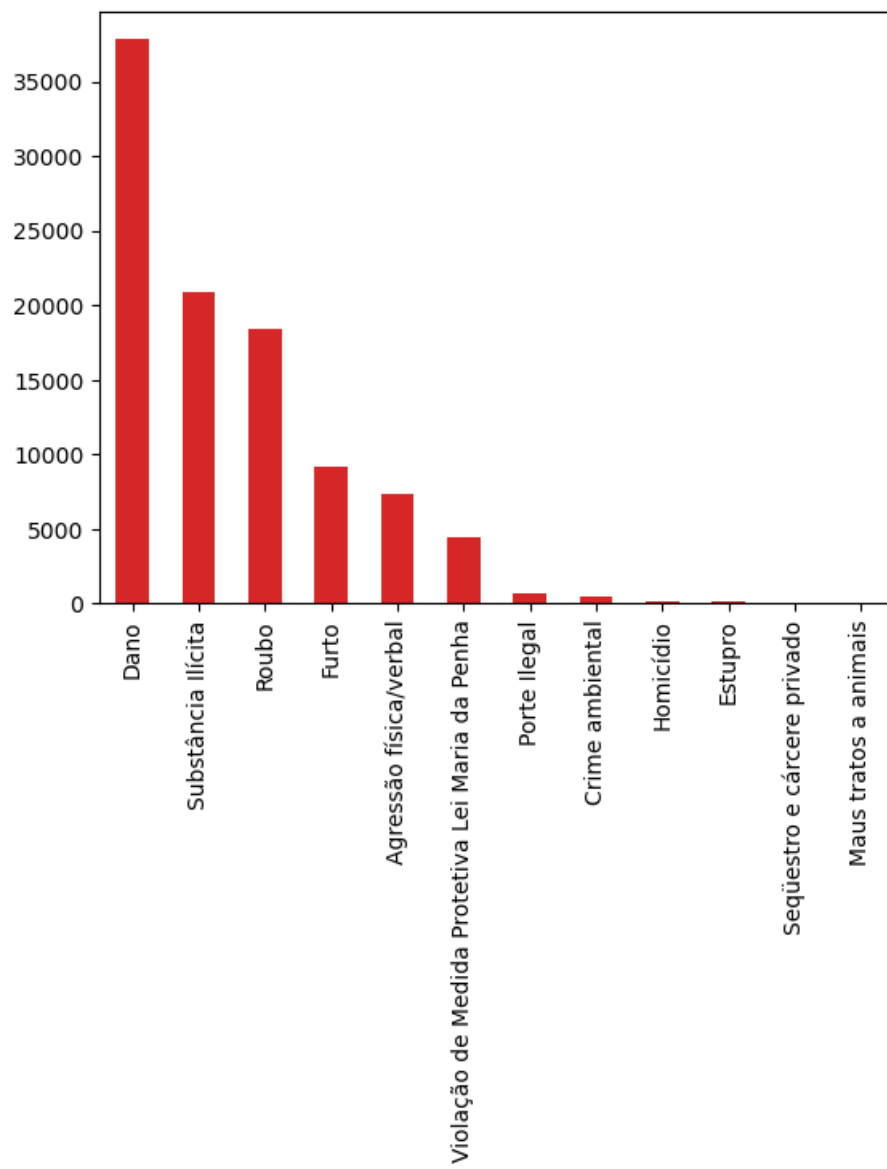


Figura 13: Quantidade total para cada tipo de crime

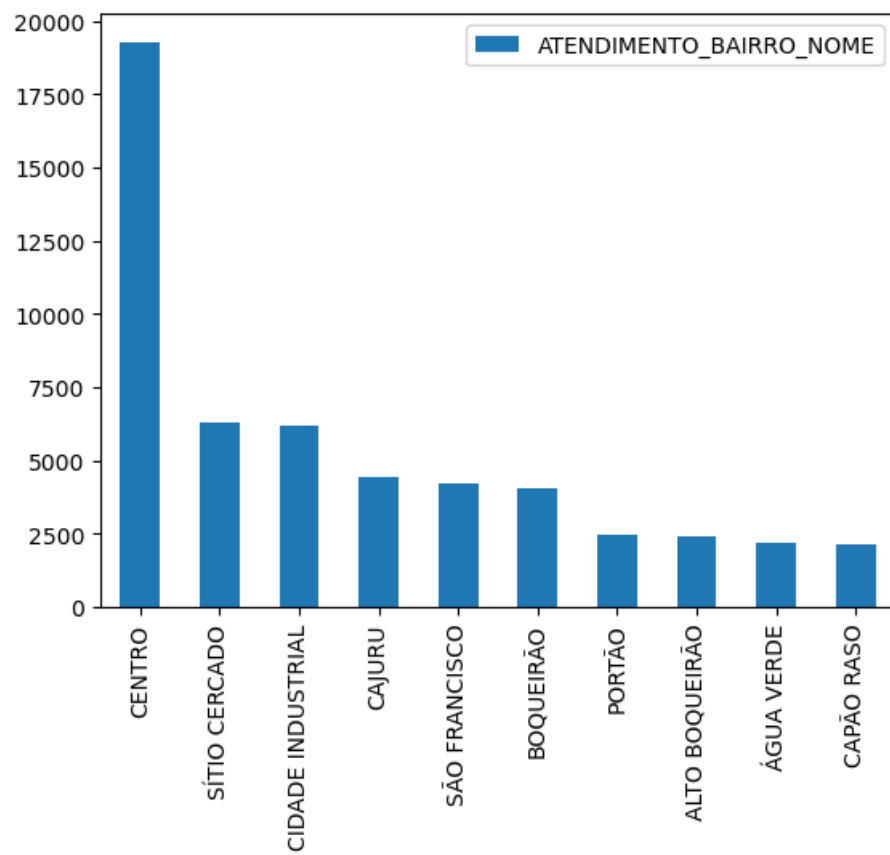


Figura 14: Quantidade de crimes por bairro

	ATENDIMENTO_BAIRRO_NOME	NATUREZA1_DESCRICAO	TOTAL
270	CENTRO	Substância Ilícita	7057
262	CENTRO	Dano	4916
268	CENTRO	Roubo	3806
285	CIDADE INDUSTRIAL	Dano	2868
617	SÍTIO CERCADO	Dano	2420
582	SÃO FRANCISCO	Substância Ilícita	2227
264	CENTRO	Furto	1991
185	CAJURU	Dano	1898
149	BOQUEIRÃO	Dano	1497
623	SÍTIO CERCADO	Roubo	1393
575	SÃO FRANCISCO	Dano	1275
260	CENTRO	Agressão física/verbal	1238
625	SÍTIO CERCADO	Substância Ilícita	1132
155	BOQUEIRÃO	Roubo	1026
293	CIDADE INDUSTRIAL	Substância Ilícita	1002
489	PORTÃO	Dano	993
18	ALTO BOQUEIRÃO	Dano	958
646	TATUQUARA	Dano	930
437	NOVO MUNDO	Dano	896
714	ÁGUA VERDE	Dano	876

Figura 15: Quantidade das 20 maiores ocorrências agrupadas por crime e bairro

Após a análise inicial, foram selecionados os 4 crimes com maior quantidade de entradas (dano, substâncias ilícitas, roubos e furtos) e análises de séries temporais separados por dia da semana, mês e ano foram realizadas:

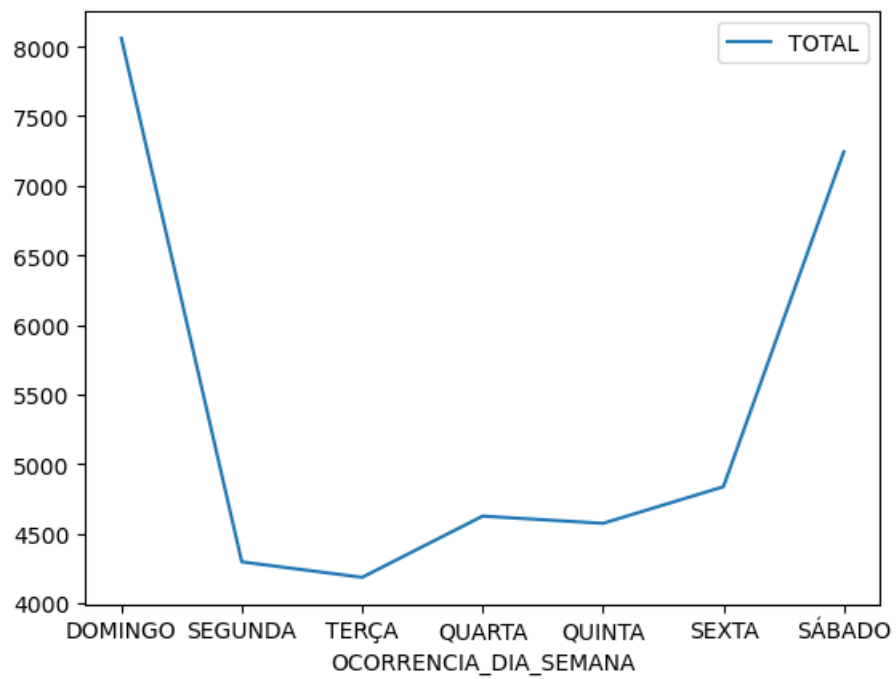


Figura 16: Quantidade de crimes relacionados a dano por dia da semana

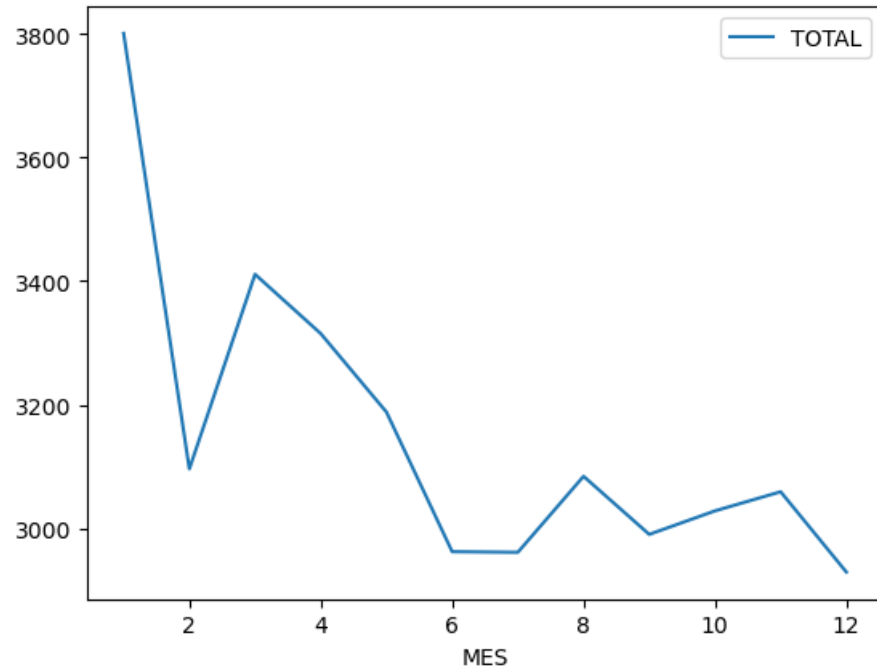


Figura 17: Quantidade de crimes relacionados a dano por mês

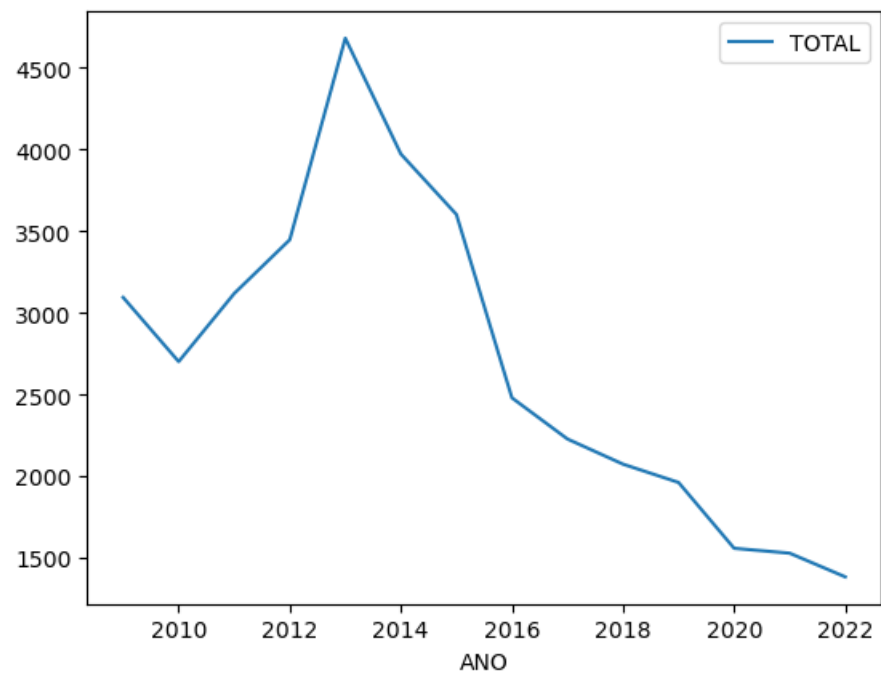


Figura 18: Quantidade de crimes relacionados a dano por ano

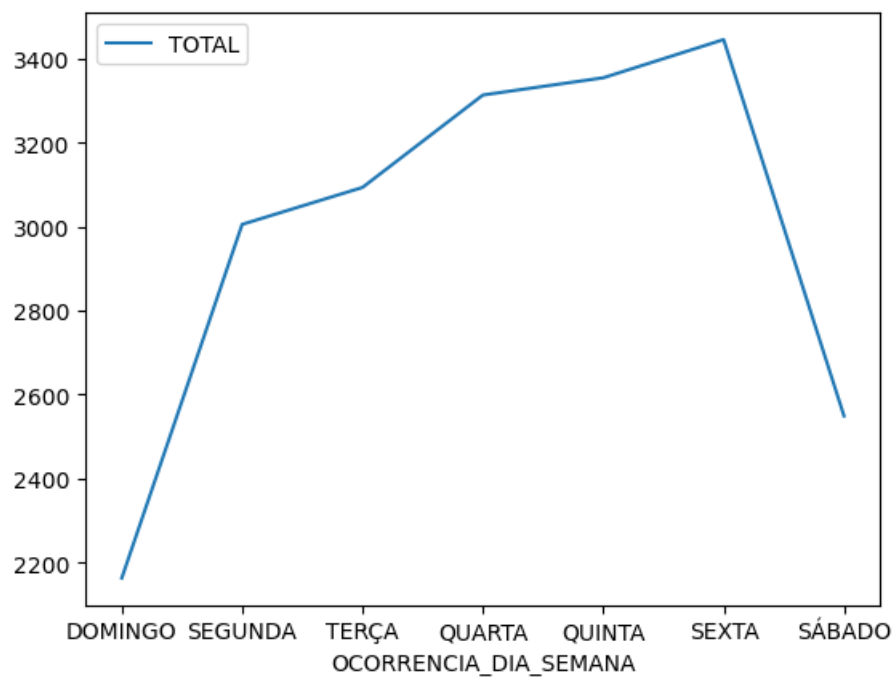


Figura 19: Quantidade de crimes relacionados a substâncias ilícitas por dia da semana

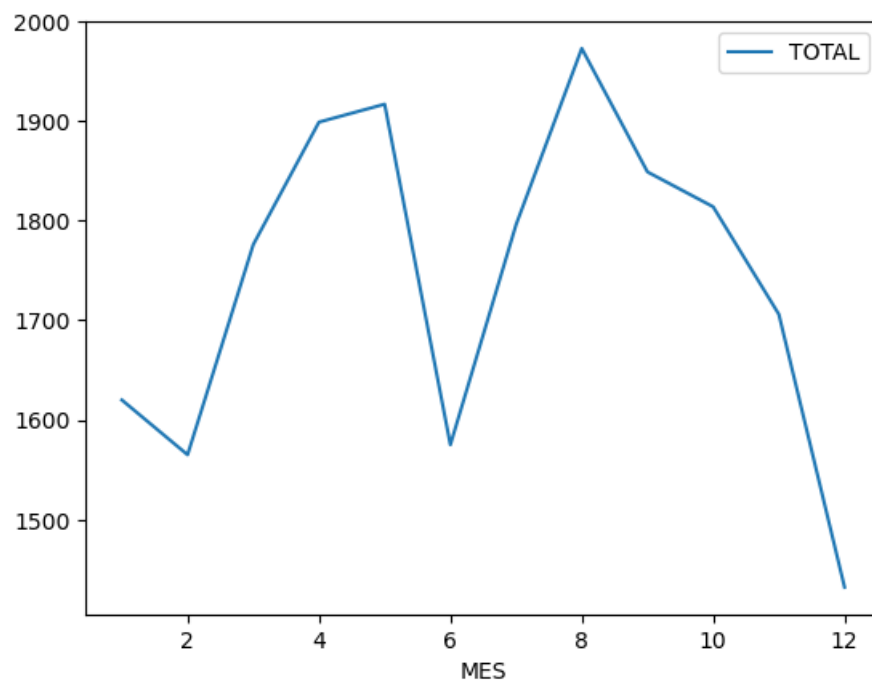


Figura 20: Quantidade de crimes relacionados a substâncias ilícitas por mês

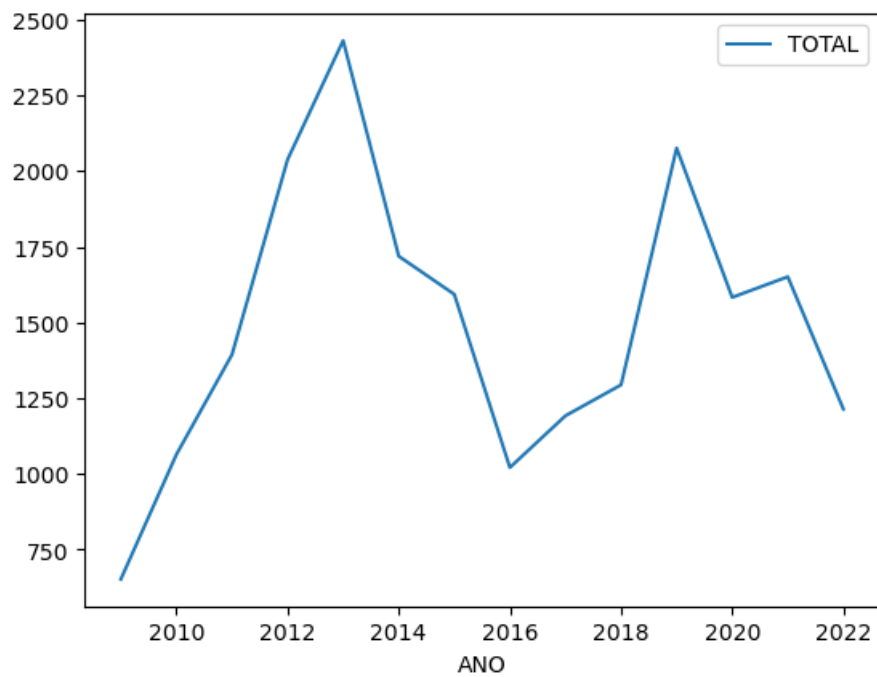


Figura 21: Quantidade de crimes relacionados a substâncias ilícitas por ano

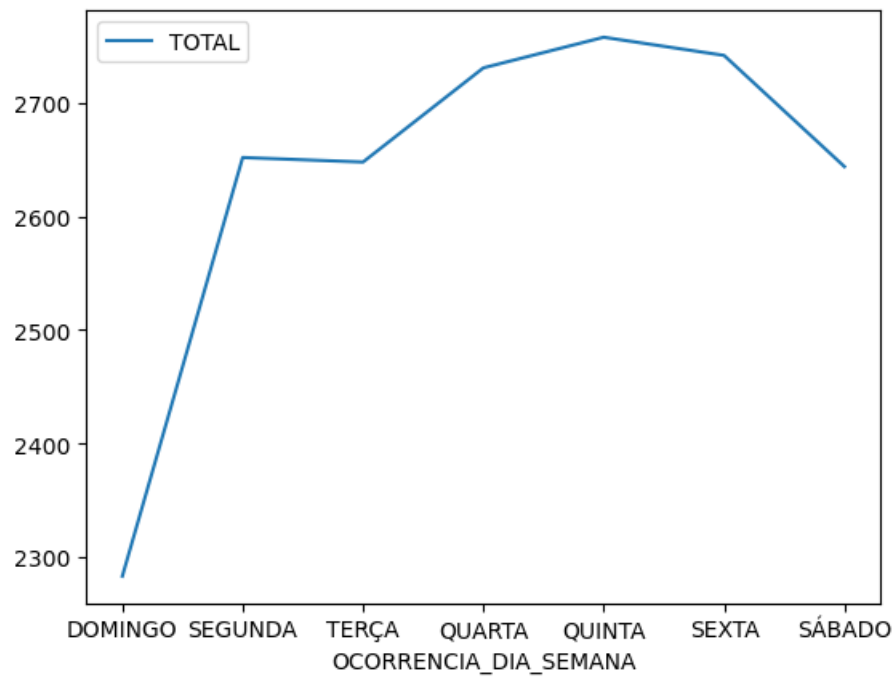


Figura 22: Quantidade de crimes relacionados a roubo por dia da semana

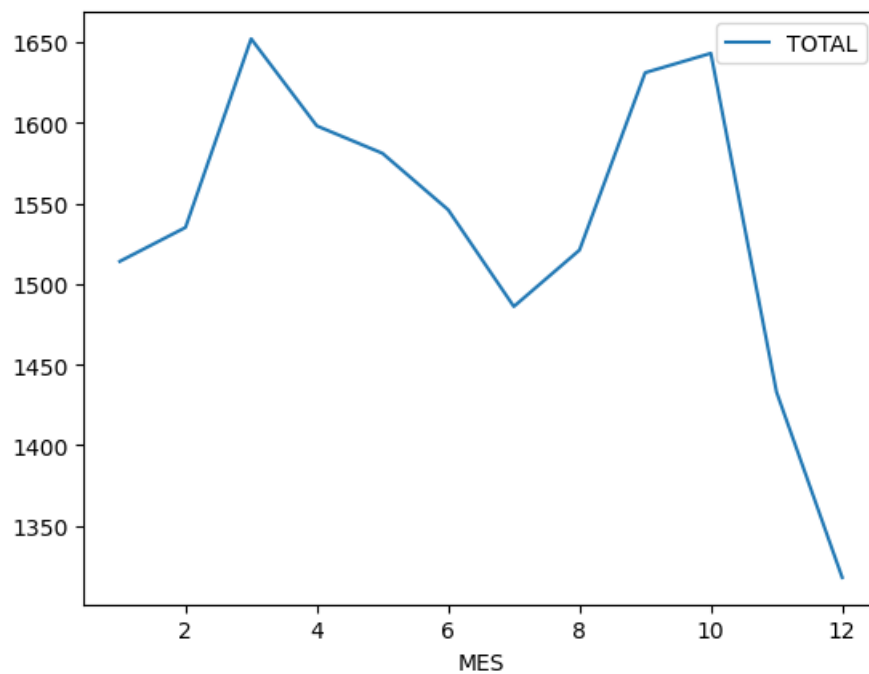


Figura 23: Quantidade de crimes relacionados a roubo por mês

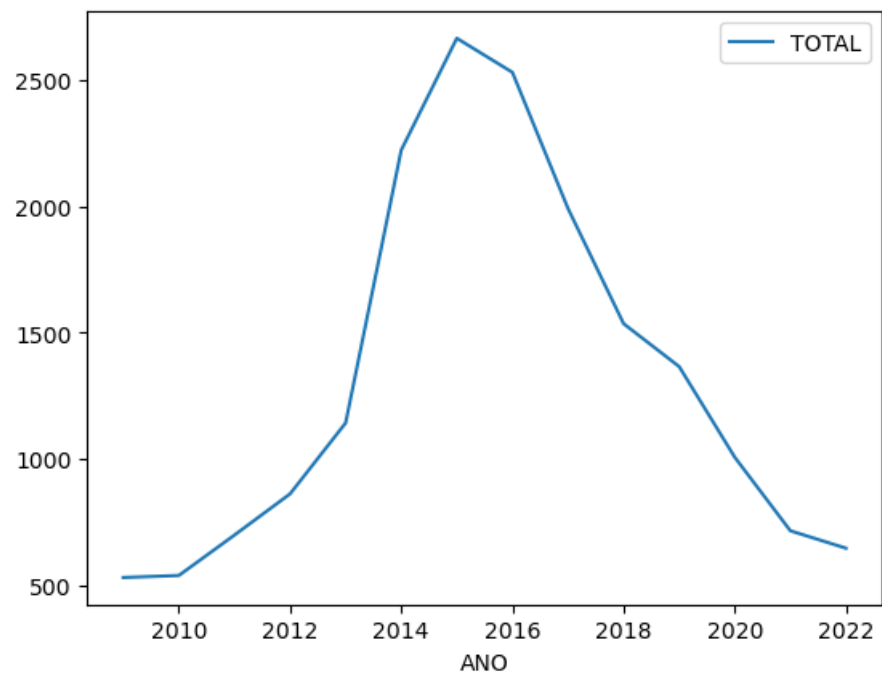


Figura 24: Quantidade de crimes relacionados a roubo por ano

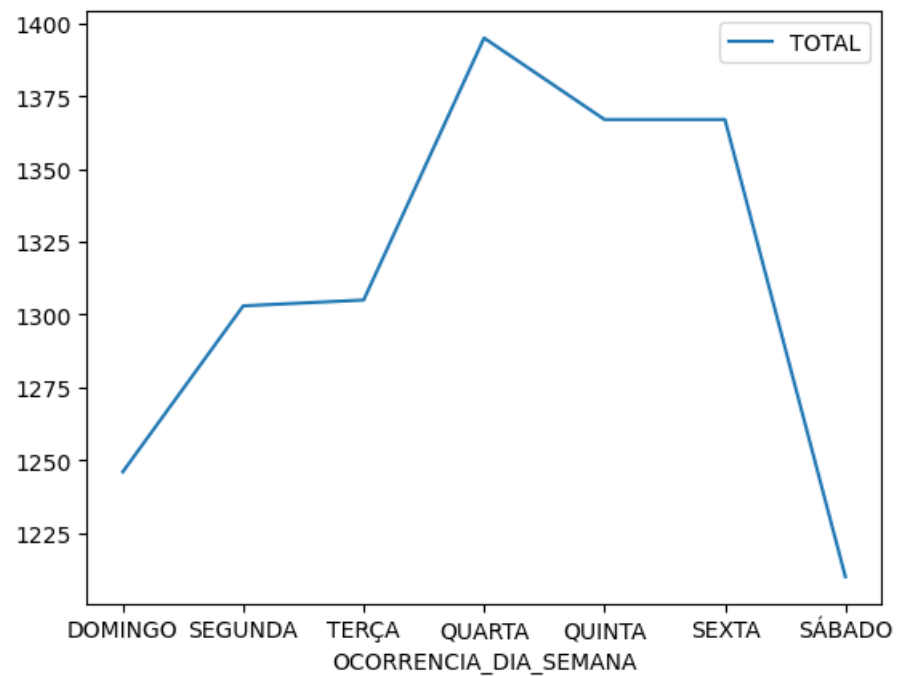


Figura 25: Quantidade de crimes relacionados a furto por dia da semana

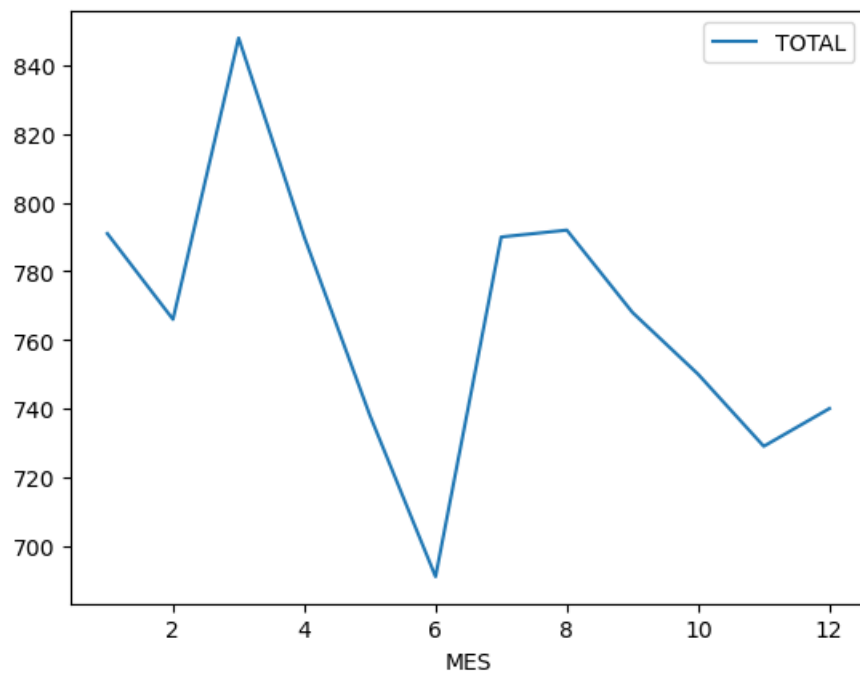


Figura 26: Quantidade de crimes relacionados a furto por mês

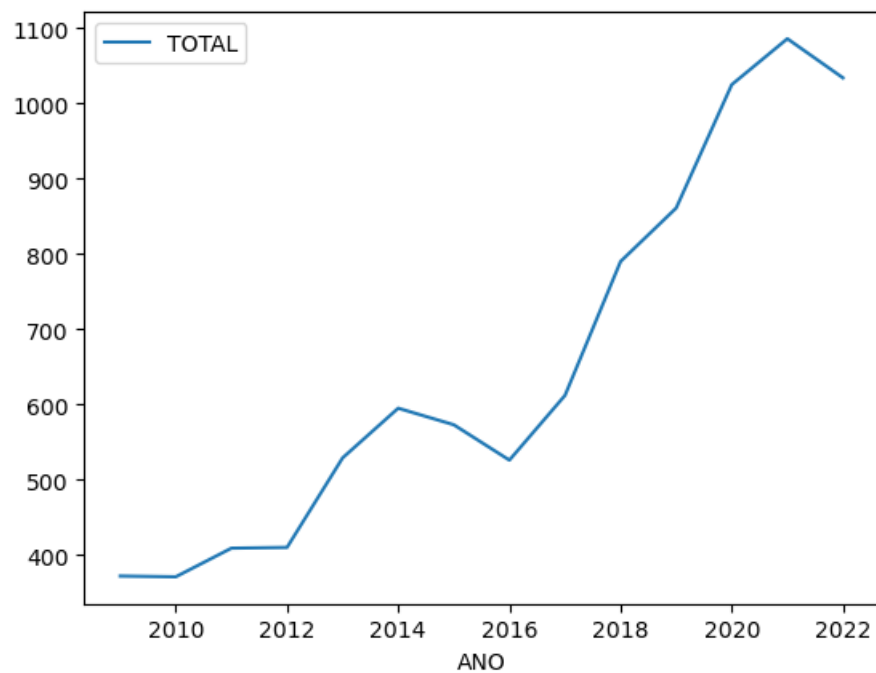


Figura 27: Quantidade de crimes relacionados a furto por ano

Através da biblioteca KModes para clusterização de dados categóricos, foram identificados 3 agrupamentos. Os grupos encontrados são descritos a seguir:

SEXTA	14081
QUINTA	6495
QUARTA	6415
TERÇA	6080
SÁBADO	3178
SEGUNDA	3003
DOMINGO	2163

Figura 28: Quantidade de ocorrências por dia da semana para o primeiro grupo

Substância Ilícita	20920
Furto	5623
Dano	4652
Agressão física/verbal	4100
Roubo	2742
Violação de Medida Protetiva Lei Maria da Penha	2553
Porte Ilegal	406
Crime ambiental	235
Homicídio	59
Seqüestro e cárcere privado	48
Estupro	40
Maus tratos a animais	37

Figura 29: Quantidade de crimes por tipo para o primeiro grupo

DOMINGO	11446
SÁBADO	10104
QUARTA	4664
QUINTA	4618
TERÇA	4239
SEXTA	186
SEGUNDA	131

Figura 30: Quantidade de ocorrências por dia da semana para o segundo grupo

Dano	29006
Agressão física/verbal	2332
Furto	2267
Violação de Medida Protetiva Lei Maria da Penha	1270
Porte Ilegal	159
Roubo	126
Crime ambiental	120
Estupro	34
Homicídio	34
Maus tratos a animais	25
Seqüestro e cárcere privado	15

Figura 31: Quantidade de crimes por tipo para o segundo grupo

SEGUNDA	9783
QUINTA	2758
QUARTA	2731
TERÇA	2648
SÁBADO	2595
DOMINGO	2206

Figura 32: Quantidade de ocorrências por dia da semana para o terceiro grupo

Roubo	15590
Dano	4166
Furto	1303
Agressão física/verbal	915
Violação de Medida Protetiva Lei Maria da Penha	565
Porte Ilegal	72
Crime ambiental	59
Estupro	18
Seqüestro e cárcere privado	13
Homicídio	9
Maus tratos a animais	9
Substância Ilícita	2

Figura 33: Quantidade de crimes por tipo para o terceiro grupo

Percebe-se que os três grupos foram agrupados de acordo com o dia da semana das ocorrências. Sendo o primeiro grupo mais representativo dos crimes cometidos nas sextas-feiras, o segundo grupo dos crimes cometidos nos fins de semana e o terceiro grupo para os crimes cometidos nas segundas-feiras. Para os demais dias, o número de ocorrência ficou distribuído para cada grupo, tendo mais entradas no primeiro grupo, em seguida no segundo, e então com menos entradas no terceiro.

Para o primeiro grupo, percebe-se que há uma quantidade relativamente maior para os crimes envolvendo substâncias ilícitas comparado com os outros tipos de crime. Podemos associar que o uso de substâncias ilícitas ocorre com mais frequência nas sextas-feiras. Para o segundo grupo, vemos mais indicações de crimes relacionados a danos, que ocorrem mais frequentemente aos fins de semana. O terceiro apresenta uma quantidade maior de crimes relacionados a roubo em relação aos outros tipos de crime, e indica uma relação com a ocorrência com as segundas-feiras.

2.2.2 DATASUS

Como primeira análise os óbitos foram agrupados por mês de ocorrência buscando procurar algum padrão dentro das ocorrências durante o ano e os dados encontrados foram:

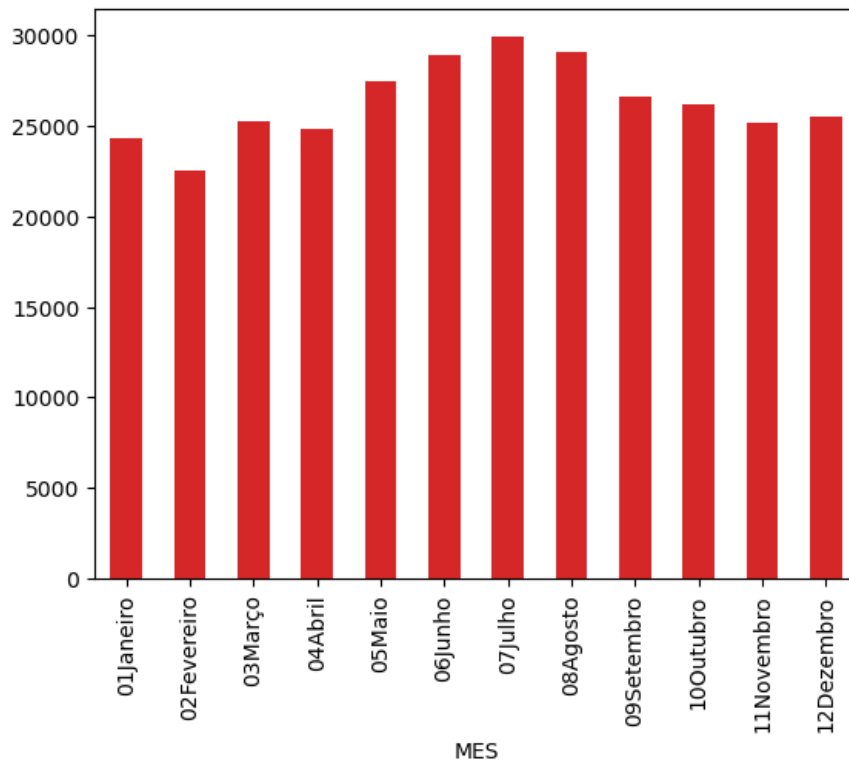


Figura 34: Óbitos por Mês de ocorrência

Fonte: Autores.

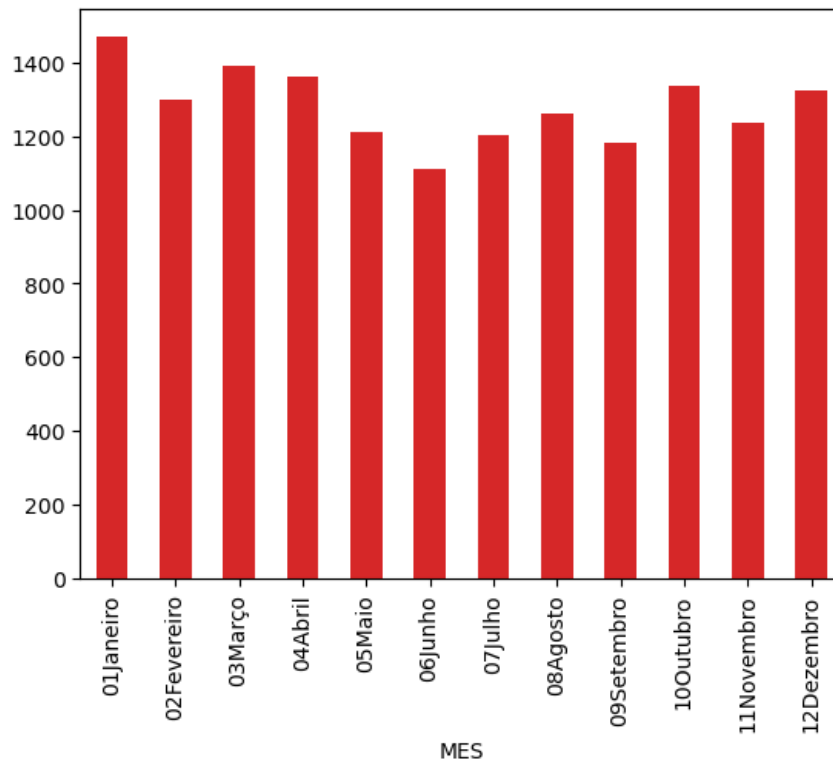


Figura 35: Óbitos por Agressão por Mês de ocorrência

Fonte: Autores.

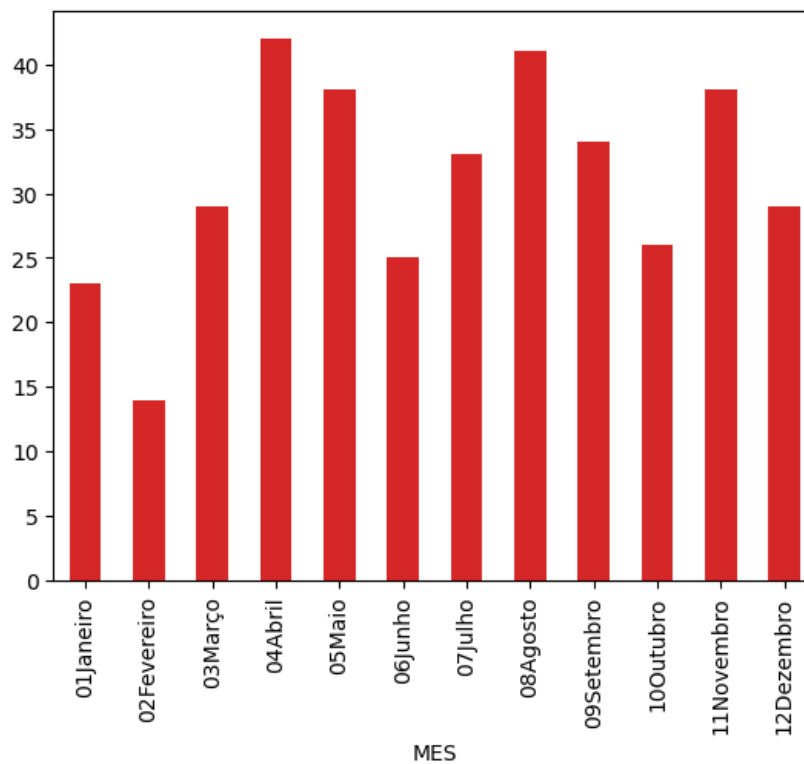


Figura 36: Óbitos por Intervenção Legal por Mês de ocorrência

Fonte: Autores.

O mesmo tipo de agrupamento foi feito também utilizando os anos de ocorrência para observar se algum padrão pode ser observado:

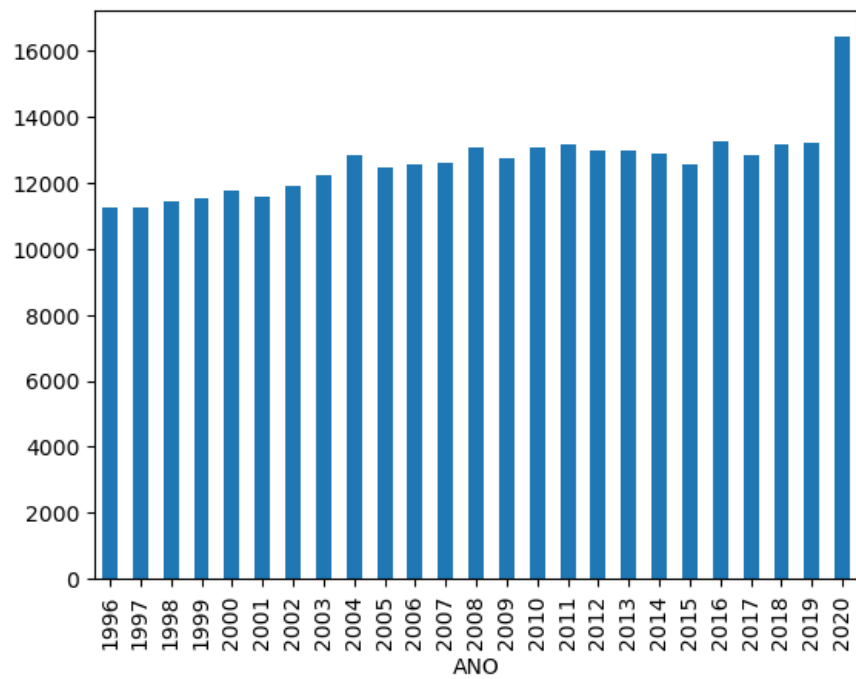


Figura 37: Óbitos por Ano de ocorrência

Fonte: Autores.

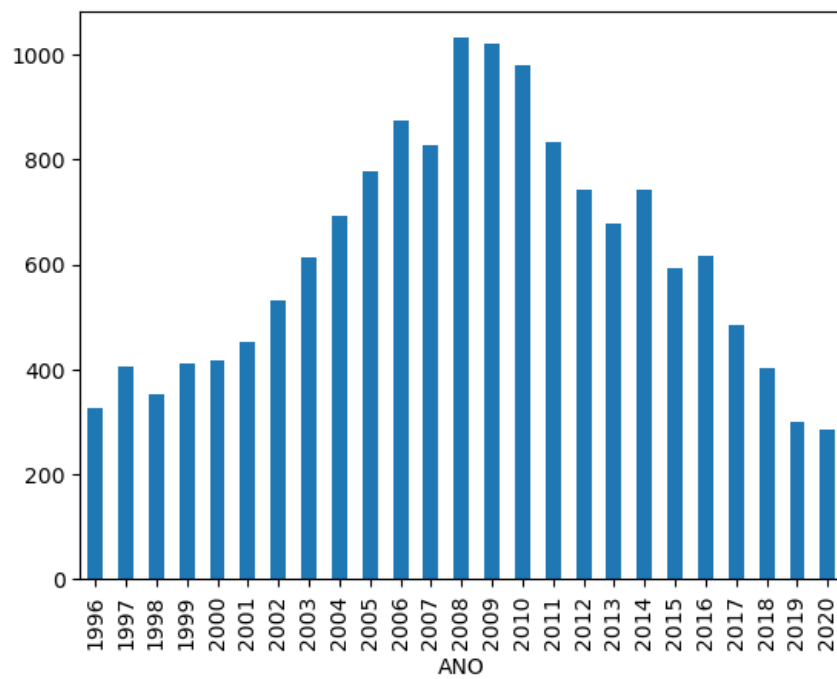


Figura 38: Óbitos por Agressão por Ano de ocorrência

Fonte: Autores.

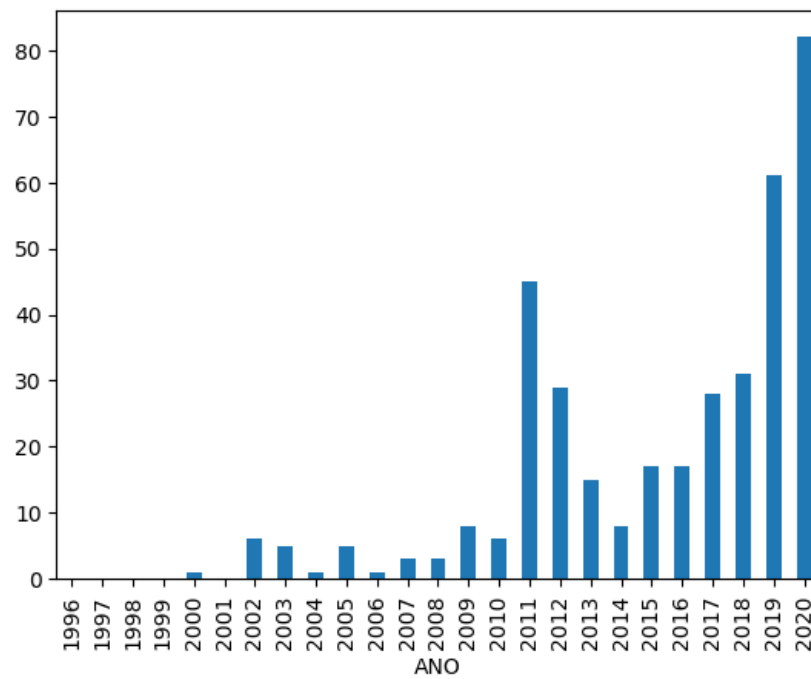


Figura 39: Óbitos por Intervenção Legal por Ano de ocorrência

Fonte: Autores.

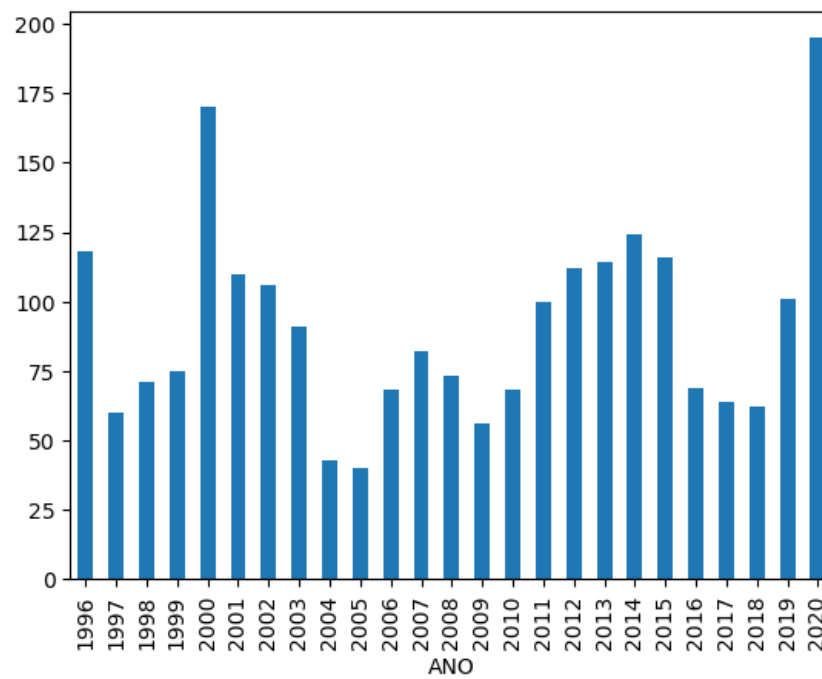


Figura 40: Óbitos Indeterminados por Ano de ocorrência

Fonte: Autores.

2.2.3 IPEA (HOMICÍDIOS)

Os dados do IPEA apresentam o número de homicídios entre 2000 e 2019 para a cidade de Curitiba. A média é de 562 homicídios por ano, sendo o menor valor 319 e o maior 778.

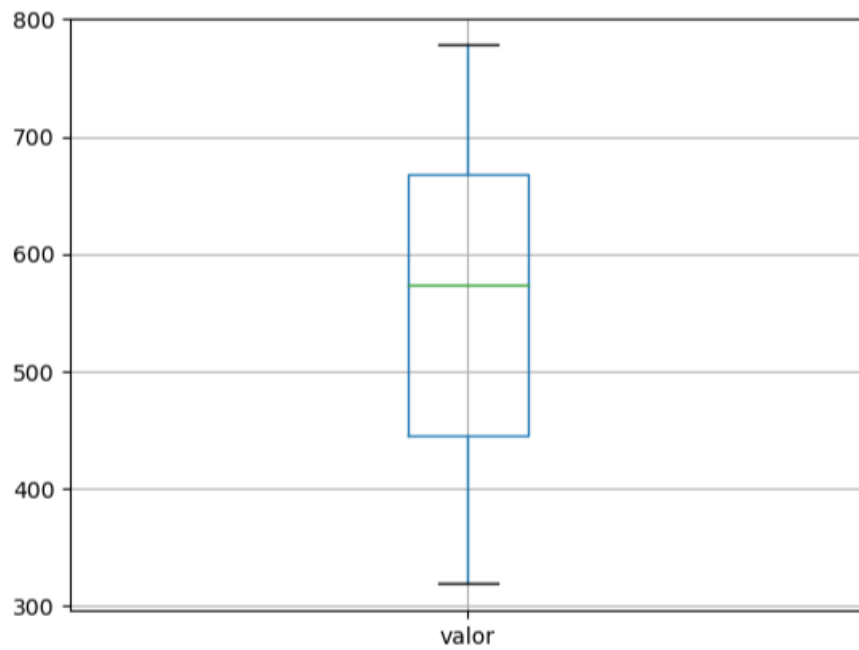


Figura 41: Resumo da quantidade de homicídios em Curitiba

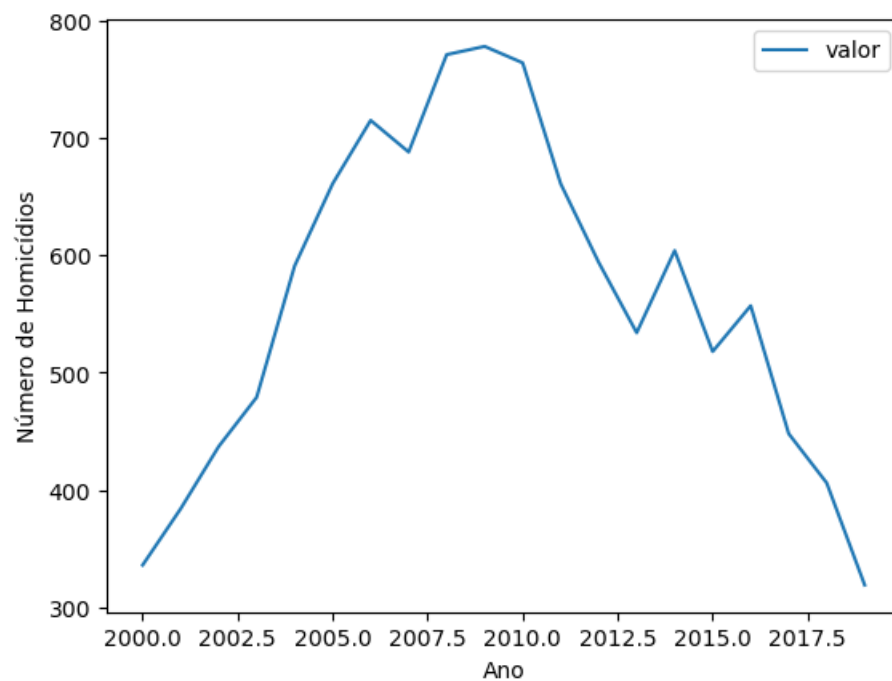


Figura 42: Homicídios em Curitiba por ano

2.2.4 IPEA (MORTES VIOLENTAS POR CAUSA INDETERMINADA)

Os dados do IPEA apresentam o número de mortes violentas por causa indeterminada entre 2000 e 2019 para a cidade de Curitiba. A média é de 64 mortes violentas por causa indeterminada por ano, sendo o menor valor 32 e o maior 136.

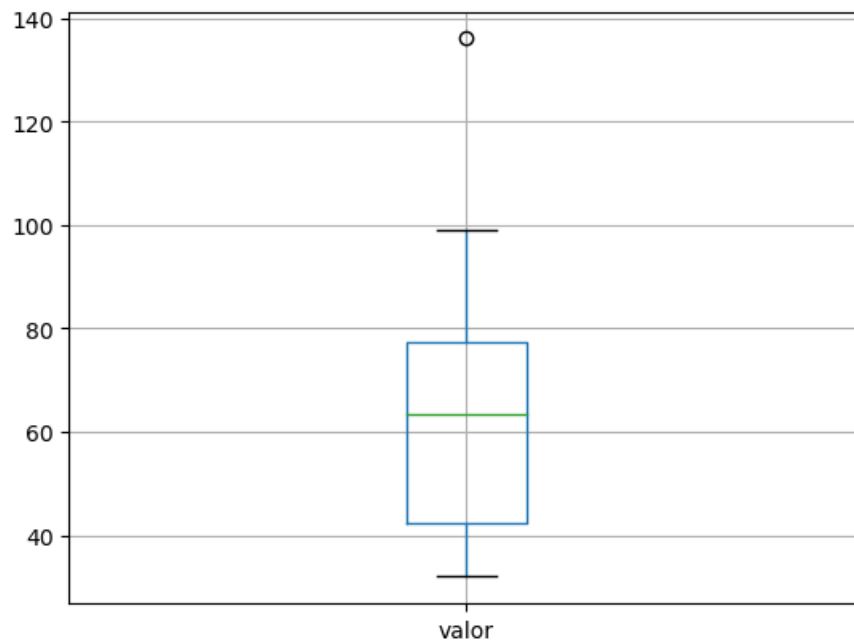


Figura 43: Resumo da quantidade de mortes violentas por causa indeterminada em Curitiba

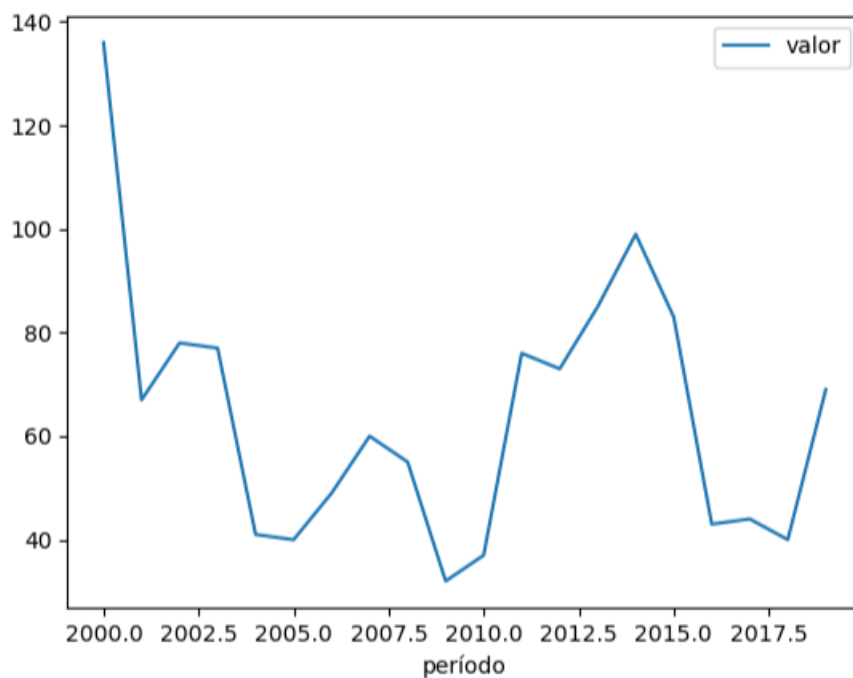


Figura 44: Mortes violentas por causa indeterminada em Curitiba por ano

2.3 IDENTIFICAÇÃO DE PADRÕES INICIAIS

2.3.1 SIGESGUARDA

2.3.2 DATASUS

Para os agrupamentos por mês pode se observar um padrão de um número maior de óbitos na cidade durante o período do inverno, entretanto observando os agrupamentos para óbitos por agressão e intervenção legal esse número maior de óbitos não se correlaciona com óbitos que podem ser atribuídos a criminalidade.

Agora nos agrupamentos por ano o número total de óbitos para o ano de 2020 se mostra claramente fora do padrão dos anos anteriores isso imagina-se ser atribuído a pandemia de COVID-19, entretanto este não é o escopo da análise.

Agora observando os agrupamentos anuais de Agressão e Intervenção Legal 2 detalhes podem ser observados, a diminuição de óbitos por agressão que ocorre desde o ano de 2011 e o aumento considerável de registros de óbitos por intervenção legal a partir de 2016.

2.3.3 IPEA

Os dados do IPEA demonstram que para Curitiba há uma tendência à redução do índice de homicídios. Os maiores índices são apresentados entre 2006 e 2010 e após este período os valores por ano têm apenas reduzido. Percebe-se que para os dados de Mortes Violentas por Causa Indeterminada houve baixos índices entre 2004 e 2010, aumento entre 2010 e 2016 e após esse período houve novamente uma redução nos números. Nos anos de 2018 e 2019 os dados voltaram a aumentar.

2.4 PERGUNTAS DE PESQUISA E EXPLORAÇÕES INICIAIS

As perguntas de pesquisa determinadas no início do projeto são:

1. Quais tipos de crime estão tendo um crescimento acelerado em quais regiões?
2. Quais tipos de crime estão tendo uma redução acelerada em quais regiões?
3. Podemos clusterizar os bairros de Curitiba de acordo com os tipos de crimes mais comuns que ocorrem nestas regiões?
4. Qual é a tendência para o futuro em relação à criminalidade em Curitiba?

5. Existiu em algum momento tipos de ocorrências que apresentaram um comportamento anômalo?
6. A pandemia influenciou os índices de criminalidade em Curitiba?
7. Existe alguma relação entre a data, hora, mês, dia da semana e as ocorrências?

Em relação às análises de tendência descritas nas perguntas 1 e 2 é possível utilizar os dados atuais para sua realização. Os dados incluem a quantidade de ocorrências e a data de quando foram registradas, sendo assim, é possível efetuar uma análise que resulta na indicação do crescimento ou redução dos índices para cada tipo de crime. Os dados também incluem infamações suficientes para realizar a análise para cada região ou ainda bairro de Curitiba, o que permite responder a terceira pergunta. Apesar disso, realizar a análise de crimes mais comuns por bairro pode não resultar em uma boa análise já que é possível que os crimes mais comuns sejam iguais para todos os bairros. Seria importante revisar esta pergunta futuramente.

A quarta pergunta também poderá ser respondida utilizando os conjuntos de dados atual, e através de gráficos e modelos de regressão será possível estimar as tendências para o futuro de Curitiba, pelo menos para um curto período de tempo. Mesmo assim, realizar este tipo de previsão pode não resultar em um modelo confiável, já que a tendência de criminalidade precisa levar em consideração múltiplas variáveis, muitas delas indetectáveis. Será possível fazer algumas previsões para poucos anos no futuro já que para certos crimes há uma clara tendência a crescimento ou decrescimento, mesmo assim, eventos inesperados poderão influenciar nos índices. Esta pergunta também terá que ser revisada, ou pelo menos experimentos deverão ser realizados para comprovar a eficiência ou ineficiência de modelos preditivos para estes tipos de dados.

Para a quinta pergunta, sobre anomalias, é possível percebê-las realizando a análise de *outliers* para as séries temporais ou até mesmo por bairros para verificar se alguma ocorrência foge da norma. Os dados são adequados para realizar este tipo de análise já que podemos utilizar o nome da ocorrência, o bairro e a data e hora para verificar se um crime foi cometido em um bairro ou data incomum.

A pergunta 6 poderá ser respondida também utilizando técnicas de detecção de *outliers* além de outras ferramentas. Os dados da SiGesGuarda apresentam ocorrências registradas também nos anos da pandemia (entre 2020 e 2022). Será possível comparar os dados de épocas anteriores com as ocorrências desses anos para verificar se a pandemia, de fato, influenciou no número de ocorrências.

A última pergunta foi explorada em alguns gráficos presentes neste documento indicando que os dados são adequados para responder esse tipo de pergunta. Alguns gráficos demonstraram que certas ocorrências acontecem mais nos fins de semana que nos demais dias e outros demonstraram que em certos meses há um rápido crescimento do número de ocorrências de certos tipos. Como a fonte de dados da SiGesGuarda apresenta a data, hora e também o dia da semana de quando foram registradas as ocorrências, conclui-se que é possível realizar todas as análises deste tipo. Será necessário, porém, filtrar a base de dados para retirar informações apenas de ocorrências relevantes.

2.5 CONCLUSÃO E PRÓXIMOS PASSOS

A partir das análises e observações realizadas nos diferentes conjuntos de dados escolhidos para estudo, pode-se trabalhar cruzando os dados advindos de fontes diferentes para poder responder as questões propostas inicialmente no trabalho para então desenvolver visualizações para demonstrar os resultados encontrados.

REFERÊNCIAS

GeoPandas Development Team. **GeoPandas: Python tools for geographic data**. 2021. Disponível em: <<https://geopandas.org/>>.

IPEA. **IPEA Atlas da Violência**. 2019. Disponível em: <<https://www.ipea.gov.br/atlasviolencia/>>.

IPPUC. **Instituto de Pesquisa e Planejamento Urbano de Curitiba - Dados Geográficos**. 2018. Acesso em 15 abril 2023. Disponível em: <<http://ippuc.org.br/geodownloads/geo.htm>>.

KLUYVER, T. et al. Jupyter notebooks – a publishing format for reproducible computational workflows. In: LOIZIDES, F.; SCHMIDT, B. (Ed.). **Positioning and Power in Academic Publishing: Players, Agents and Agendas**. [S.l.], 2016. p. 87 – 90.

Matplotlib Development Team. **Matplotlib: A 2D graphics environment**. 2021. Disponível em: <<https://matplotlib.org/>>.

Pandas Development Team. **Pandas: Powerful data analysis tools for Python**. 2021. Disponível em: <<https://pandas.pydata.org/>>.

Prefeitura Municipal de Curitiba. **Dicionário de Dados do Sistema Integrado de Gestão de Guarda Municipal de Curitiba**. 2015. Disponível em: <<https://encurtador.com.br/mEHJV>>.

Prefeitura Municipal de Curitiba. **Casos de COVID-19 em Curitiba**. 2023. Disponível em: <<https://www.curitiba.pr.gov.br/dadosabertos/busca/?grupo=16>>.

Prefeitura Municipal de Curitiba. **Portal de Dados Abertos de Curitiba**. acesso em 16 abril 2023. Disponível em: <<https://www.curitiba.pr.gov.br/dadosabertos/busca/>>.

ROSSUM, G. V.; DRAKE, F. L. **Python 3 Reference Manual**. Scotts Valley, CA: CreateSpace, 2009. ISBN 1441412697.

SAÚDE, M. da. **DataSUS TABNET**. 2020. Disponível em: <<https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>>.

TRÂNSITO, D. S. e. **Base de Dados SiGesGuarda**. 2022. Disponível em: <<https://www.curitiba.pr.gov.br/dadosabertos/busca/?termo=guarda>>.