

# Atividade 02 - Análise Exploratória

## Tema

Análise da criminalidade em Curitiba - Evolução temporal, distribuição espacial e correlação com os indicadores de renda, segurança e desenvolvimento das regionais da cidade de Curitiba.

## Equipe

**Nome:** CSI Curitibanos

**Repositório GitLab da equipe:** <https://gitlab.com/thiagorbernardo/IntroCD>

### Membros:

- |  |                 |                           |
|--|-----------------|---------------------------|
| • João Felipe Sarggin Machado,<br><a href="mailto:joaomachado@alunos.utfpr.edu.br">joaomachado@alunos.utfpr.edu.br</a> ,               | 1905023,<br>EC, | jfsmachado,<br>UTFPR      |
| • João Lucas Mizuguchi Ferreira,<br><a href="mailto:joaolucasferreira@alunos.utfpr.edu.br">joaolucasferreira@alunos.utfpr.edu.br</a> , | 2135965,<br>EC, | jlmizug,<br>UTFPR         |
| • Pedro Sodré dos Santos,<br><a href="mailto:psantos.2000@alunos.utfpr.edu.br">psantos.2000@alunos.utfpr.edu.br</a> ,                  | 2126745,<br>EC, | pedro_sodre,<br>UTFPR     |
| • Thiago Ramos Bernardo,<br><a href="mailto:thiagobernardo@alunos.utfpr.edu.br">thiagobernardo@alunos.utfpr.edu.br</a> ,               | 2126753,<br>EC, | thiagorbernardo,<br>UTFPR |

Visualizando as coisas básicas do dataset, foi visto que existiam muitas (185) naturezas de crime (Roubo, Furto, Desacato, Invasão, etc.), então, foi considerado as naturezas mais relevantes e comuns, reduzindo o número de naturezas de 185 para 41, para serem analisadas.

Após essas considerações, as análises foram divididas em dois caminhos: de um lado, teve uma busca para entender a relação entre os índices dos bairros com os tipos de crimes realizados, e para isso, foi feito uma junção do dataset dos índices dos bairros com os crimes através da coluna que continha o nome do bairro; do outro lado, foi feito uma análise buscando analisar a relação entre dias e horários, com os crimes realizados, dessa forma a análise foi dividida em 2 jupyter notebooks.

## Cobertura e distribuição dos dados

Para verificar a distribuição dos dados, antes de tudo foi verificado os crimes mais cometidos dentro da cidade, verificando o top 10 crimes mais cometidos em Curitiba, que pode ser visto pela tabela abaixo:

| NATUREZA1 | DESCRICAO                                       | TOTAL_POR_CRIME |
|-----------|---|-----------------|
|           | Roubo   | 17925           |
|           | Furto   | 8346            |
|           | Ameaça  | 7762            |
|           | Agressão física/verbal                          | 6573            |
|           | Violação de Medida Protetiva Lei Maria da Penha | 3681            |
|           | Risco de acidente / à vida                      | 2755            |
|           | Alagamento                                      | 2396            |
|           | Atos obscenos/libidinosos                       | 1800            |
|           | Disparo de arma                                 | 1358            |
|           | Lesão Corporal                                  | 936             |

**Figura 3. Tabela de crimes mais cometidos**

Dessa forma, é verificado que a maior parte de crimes cometidos na cidade são de roubos e furtos, além de uma grande quantidade de ameaças e agressões.

Para entender melhor, e termos gráficos mais relevantes foi feito um agrupamento dos tipos de ocorrências, juntando roubo e furto, crimes de pudor etc, como é possível ver na tabela abaixo:

|   | NATUREZA     | TOTAL_POR_CRIME |
|---|--------------|-----------------|
| 7 | ROUBO        | 26839           |
| 0 | AGRESSAO     | 15361           |
| 6 | RISCO_VIDA   | 4382            |
| 1 | CRIME_MULHER | 3681            |
| 5 | PUDOR        | 2805            |
| 8 | SEQUESTRO    | 415             |
| 4 | MORTES       | 141             |
| 2 | ESTUPRO      | 108             |
| 3 | EXPLORACAO   | 17              |

**Figura 4. Tabela de crimes mais cometidos**

Após isso, como números totais não são bons de fazer a análise, foi feita a proporção de crimes com sua população, adicionando uma nova coluna ('Proporcao') no dataset. Como é possível ver na tabela abaixo.

|    | ATENDIMENTO_BAIRRO_NOME | POPULACAO_TOTAL | TOTAL_POR_BAIRRO | PROPORCAO |
|----|-------------------------|-----------------|------------------|-----------|
| 37 | RIVIERA                 | 289             | 2                | 0.006920  |
| 30 | LAMENHA PEQUENA         | 1056            | 10               | 0.009470  |
| 18 | CASCATINHA              | 2161            | 36               | 0.016659  |
| 19 | CAXIMBA                 | 2522            | 74               | 0.029342  |
| 28 | HUGO LANGE              | 3392            | 51               | 0.015035  |
| 10 | BOM RETIRO              | 5156            | 93               | 0.018037  |
| 29 | JARDIM SOCIAL           | 5698            | 47               | 0.008249  |
| 36 | PRADO VELHO             | 6077            | 504              | 0.082936  |
| 3  | AUGUSTA                 | 6598            | 192              | 0.029100  |
| 15 | CAMPINA DO SIQUEIRA     | 7326            | 542              | 0.073983  |
| 32 | ORLEANS                 | 8105            | 110              | 0.013572  |
| 23 | FANNY                   | 8415            | 236              | 0.028045  |
| 1  | ALTO DA RUA XV          | 8531            | 315              | 0.036924  |
| 13 | CACHOEIRA               | 9314            | 128              | 0.013743  |
| 7  | BATEL                   | 10878           | 481              | 0.044218  |
| 25 | GANCHINHO               | 11178           | 262              | 0.023439  |
| 43 | VISTA ALEGRE            | 11199           | 169              | 0.015091  |
| 26 | GUABIOTUBA              | 11461           | 228              | 0.019894  |
| 33 | PAROLIN                 | 11554           | 589              | 0.050978  |
| 42 | VILA IZABEL             | 11610           | 145              | 0.012489  |
| 40 | TINGUI                  | 12319           | 195              | 0.015829  |
| 11 | RUJATIUVINHA            | 12876           | 160              | 0.012426  |

**Figura 5. Tabela de Bairros com a população, quantidade total de crimes, e a proporção (Crimes/População)**

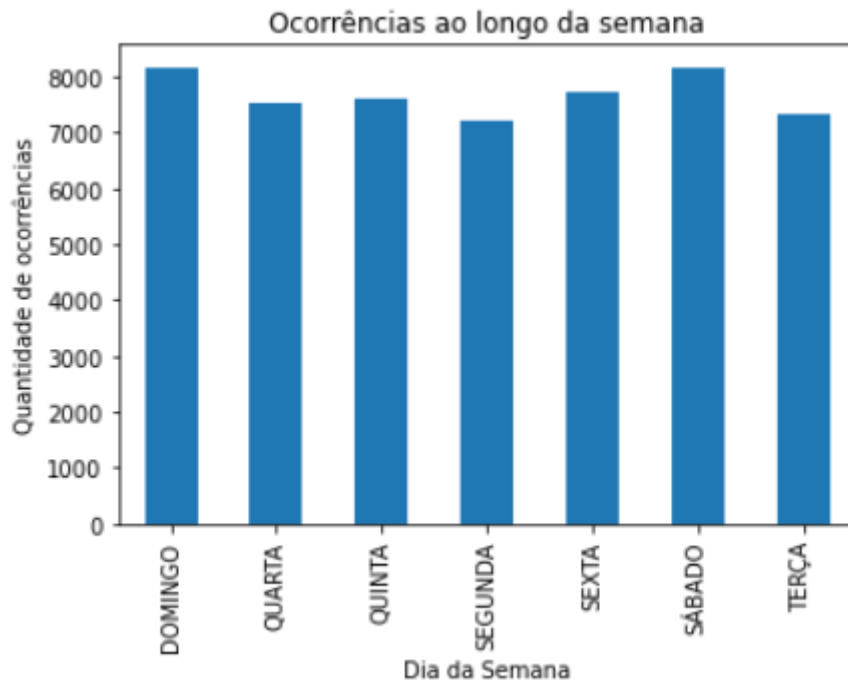
Também foi verificada a distribuição do tipo de crime com maior quantidade que é o roubo, juntamente com um dos parâmetros considerados primariamente importantes na análise, que é o rendimento mensal mediano, o que significa que pelo menos metade da população recebe menos que esse valor. Dessa forma foi-se obtido a tabela abaixo:

|      | ATENDIMENTO_BAIRRO_NOME       | POPULACAO_TOTAL | RENDIMENTO_MENSAL_MEDIANO | TOTAL_ROUBOS | PROPORCAO |
|------|-------------------------------|-----------------|---------------------------|--------------|-----------|
| 938  | TATUQUARA                     | 52780           | 510.0                     | 117          | 0.002217  |
| 478  | CAXIMBA                       | 2522            | 510.0                     | 6            | 0.002379  |
| 874  | PRADO VELHO                   | 6077            | 510.0                     | 243          | 0.039987  |
| 643  | GANCHINHO                     | 11178           | 510.0                     | 18           | 0.001610  |
| 93   | AUGUSTA                       | 6598            | 520.0                     | 23           | 0.003486  |
| 446  | CAMPO DE SANTANA              | 26657           | 580.0                     | 55           | 0.002063  |
| 989  | UBERABA                       | 72056           | 600.0                     | 122          | 0.001693  |
| 290  | BUTIATUVINHA                  | 12876           | 600.0                     | 30           | 0.002330  |
| 726  | LAMENHA PEQUENA               | 1056            | 600.0                     | 3            | 0.002841  |
| 794  | PAROLIN                       | 11554           | 600.0                     | 316          | 0.027350  |
| 849  | PINHEIRINHO                   | 50401           | 600.0                     | 524          | 0.010397  |
| 550  | CIDADE INDUSTRIAL DE CURITIBA | 172822          | 600.0                     | 623          | 0.003605  |
| 366  | CAJURU                        | 96200           | 600.0                     | 849          | 0.008825  |
| 331  | CACHOEIRA                     | 9314            | 600.0                     | 13           | 0.001396  |
| 21   | ABRANCHES                     | 13189           | 650.0                     | 46           | 0.003488  |
| 769  | ORLEANS                       | 8105            | 700.0                     | 22           | 0.002714  |
| 616  | FAZENDINHA                    | 28074           | 700.0                     | 217          | 0.007730  |
| 70   | ATUBA                         | 15935           | 700.0                     | 44           | 0.002761  |
| 819  | PILARZINHO                    | 28480           | 700.0                     | 46           | 0.001615  |
| 1052 | XAXIM                         | 57182           | 700.0                     | 582          | 0.010178  |

**Figura 6. Tabela de Bairros com a população, rendimento mensal mediano, total de roubos e a proporção (Roubo/População)**

## Identificação de Padrões Iniciais

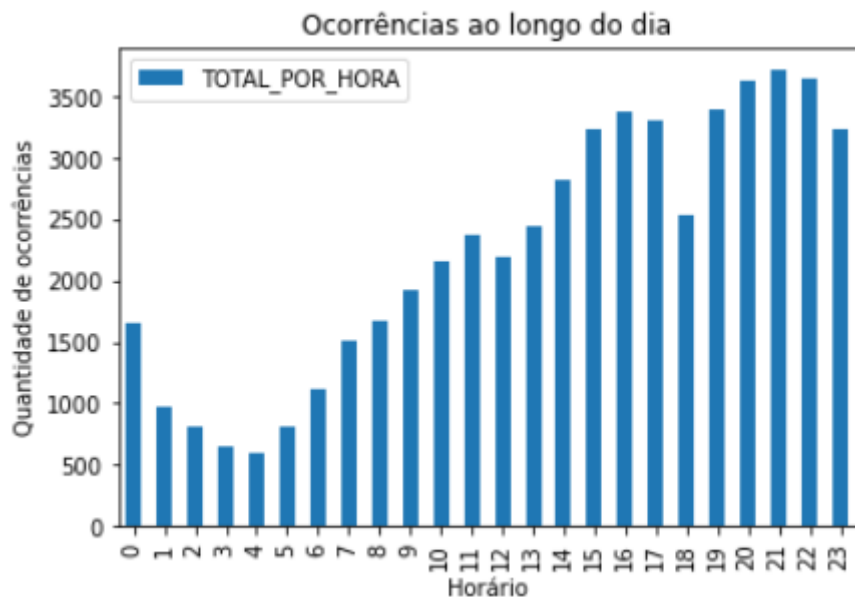
Foi tentado verificar se existia alguma relação entre os dias da semana, se haveria uma maior ocorrência de crimes em algum dia específico, dessa forma, os crimes foram agrupados pelas quantidades de dias da semana, gerando o seguinte gráfico:



**Figura 7. Gráfico de Barras mostrando a ocorrências crimes por dia da semana**

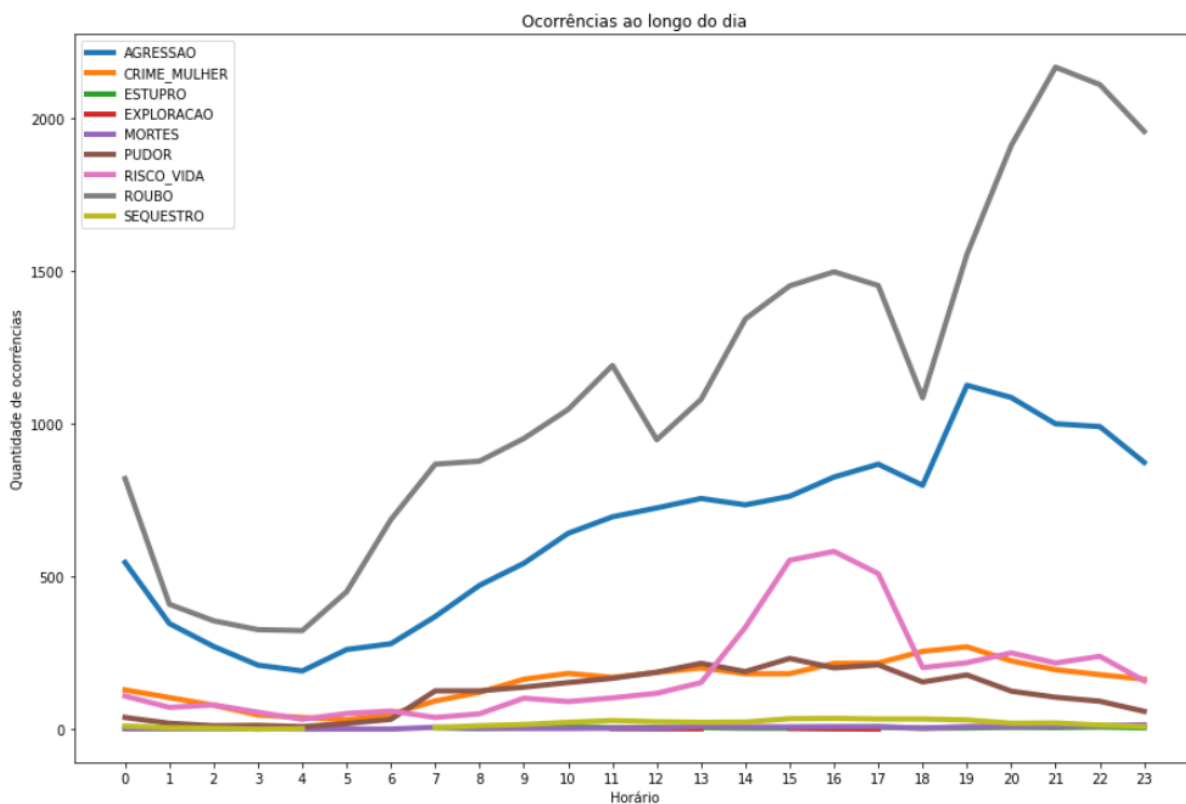
Com esse gráfico, foi possível visualizar que não há uma relação tão explícita de um período da semana que ocorre mais crimes que outros, é possível verificar que no início da semana (Segunda e Terça) são os dias que possuem menos ocorrências. Porém a diferença não chega a ser nem de 10% em relação aos outros dias.

Também foi verificado se, da mesma forma que foi feita com os dias da semana, foi feita com os horários do dia, buscando em que período do dia seria mais provável de ocorrer algum tipo de crime, no entanto, para isso, foi necessário separar os horários dos minutos, montando uma nova coluna com apenas a hora registrada para fazer a análise. E, após isso, foi verificado os seguinte gráficos de horas do dia por quantidade de crimes cometidos:

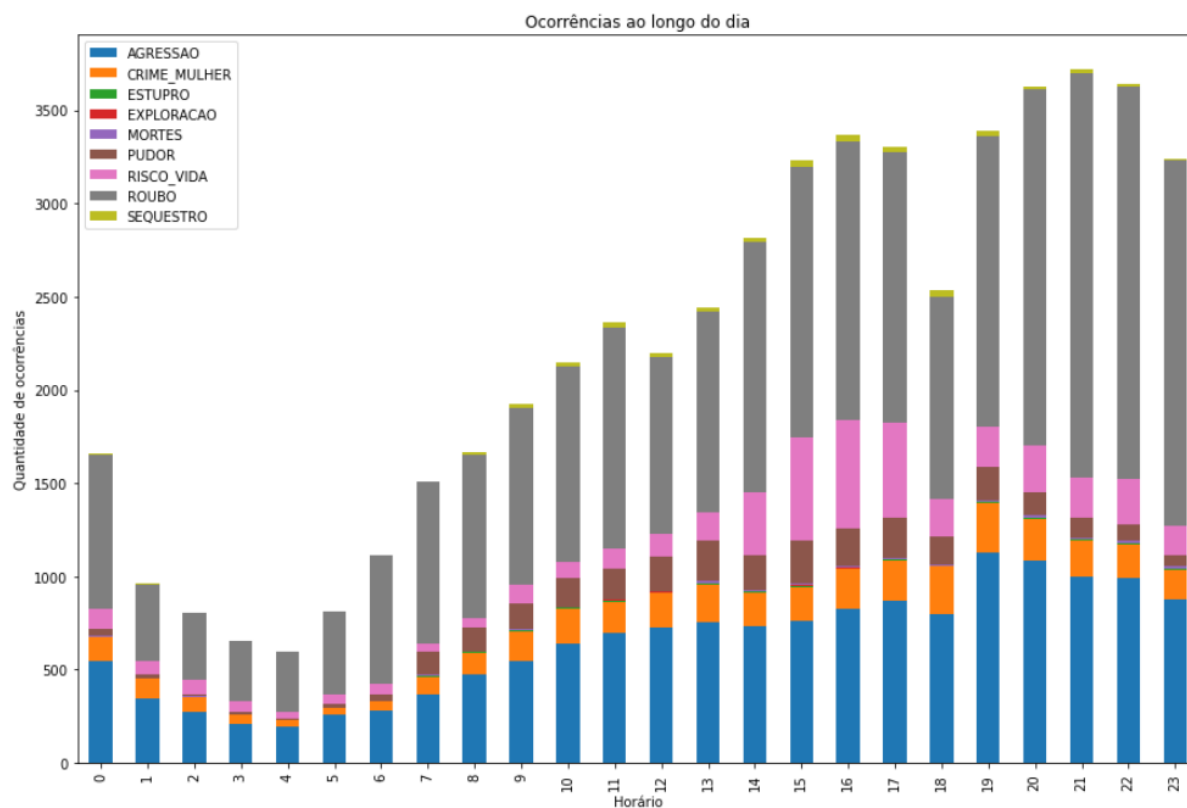


**Figura 8. Gráfico de Barras mostrando a ocorrência de crimes em cada hora**

Dessa forma, como já era de se esperar pelo senso comum, a maior quantidade de crimes ocorrem no período da noite, e há uma grande diferença do período da madrugada juntamente com início da manhã (da 00h às 6h), com o período da noite (das 19h às 23h), com mais 3,5 vezes mais crimes ocorrendo no período noturno. Para entender melhor essas ocorrências foi separado pelo tipo e plotado a quantidade de cada um.



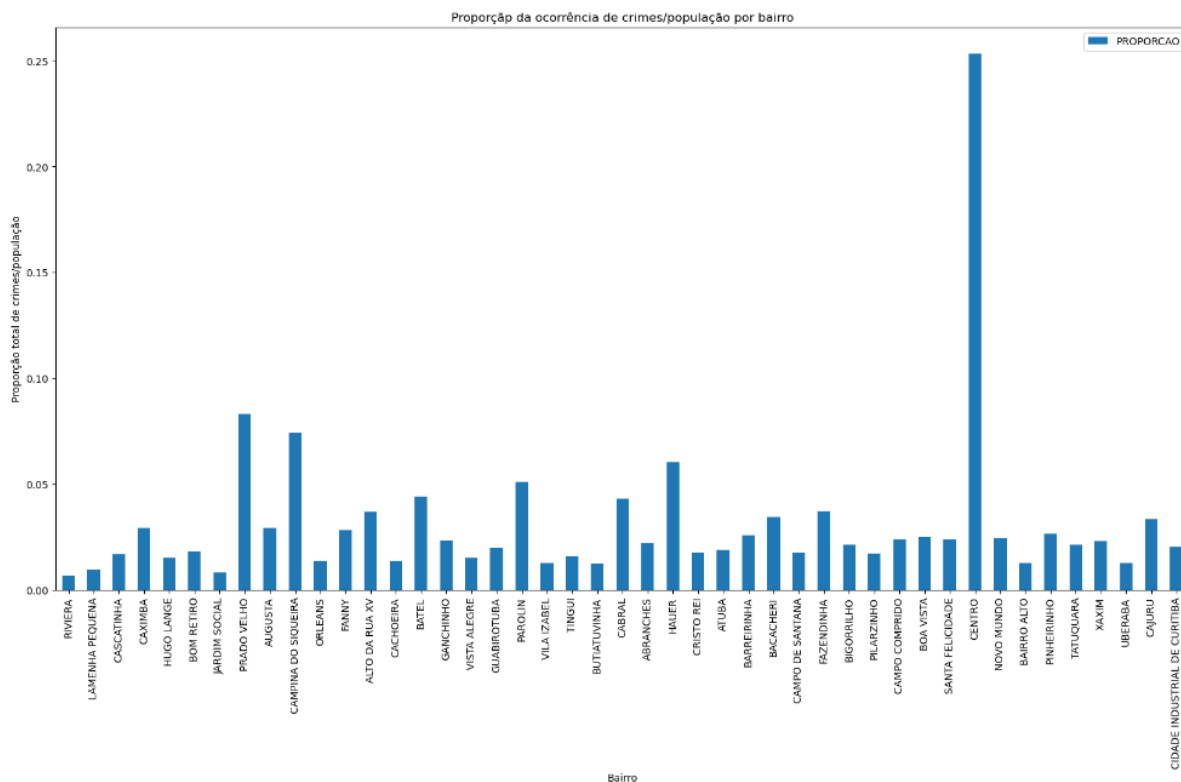
**Figura 9. Gráfico de linha mostrando a ocorrência de cada crime em cada hora**



**Figura 10. Gráfico de barra empilhado mostrando a ocorrência de cada crime em cada hora**

É possível ver uma maior quantidade de roubos, sendo o tipo de reclamação/crime mais comum independente do horário. Roubos aumentam bastante depois das 10h e tem um pico às 21h, já agressão tem um pico às 19h. Uma observação interessante é que ocorrências do tipo de risco a vida ocorrem mais entre as 15h e 17h. Porém, crimes a mulher e outras categorias são constantes durante o dia, apenas diminuindo na madrugada

Depois partindo para uma análise dos bairros para os crimes foi feito uma agregação dos bairros que gerou o seguinte gráfico:

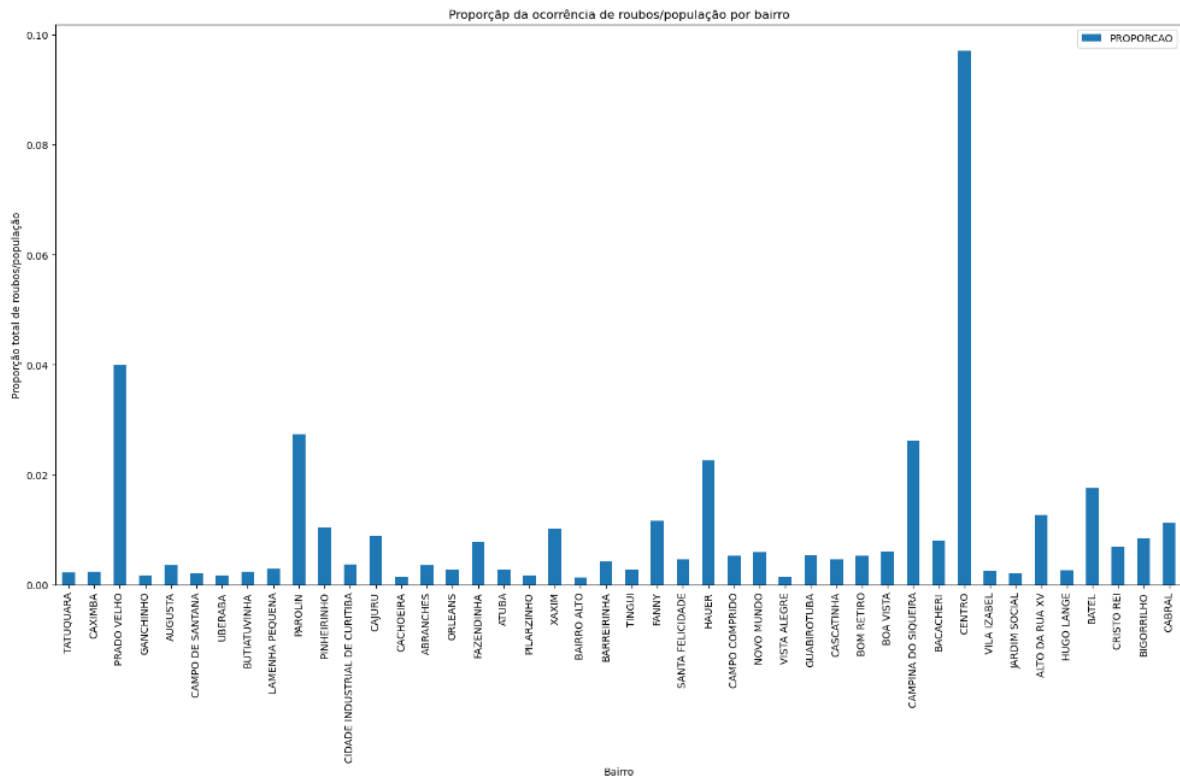


**Figura 9. Gráfico de Barras mostrando a quantidade de crimes em cada bairro.**

Como é possível observar no gráfico, o bairro com a maior quantidade proporcional de crimes é o Centro, que é o nono bairro mais populoso, já os outros 8 bairros mais populosos possuem uma quantidade menor de crimes. Com isso é possível concluir que o tamanho da população é um fator que não tem um peso grande na ocorrência de crimes. Também foi possível verificar que a área mais afetada é o centro, com cerca de 0,25 crimes por habitante, maior que o dobro da segunda maior, o que significa uma grande taxa de crimes (podemos aproximar que a cada 4 habitantes no centro, 1 foi vítima de crime).

Depois foi feita uma análise do crime mais cometido na cidade: o roubo. Como é possível observar no gráfico, o bairro com a maior quantidade de roubos é o Centro, que é o nono bairro com maior renda mensal mediano, já os outros 8 bairros com maior renda mensal mediano possuem uma quantidade menor de roubos. Com isso é possível concluir que a renda mensal mediana é um fator que não tem um peso grande na ocorrência de roubos. Também foi possível verificar que a área mais afetada é o centro, com cerca de 0,10 roubos por habitante, maior que o dobro da segunda maior, o que significa uma grande taxa de roubos (podemos aproximar que a cada 10 habitantes no centro, 1 foi vítima de roubo).



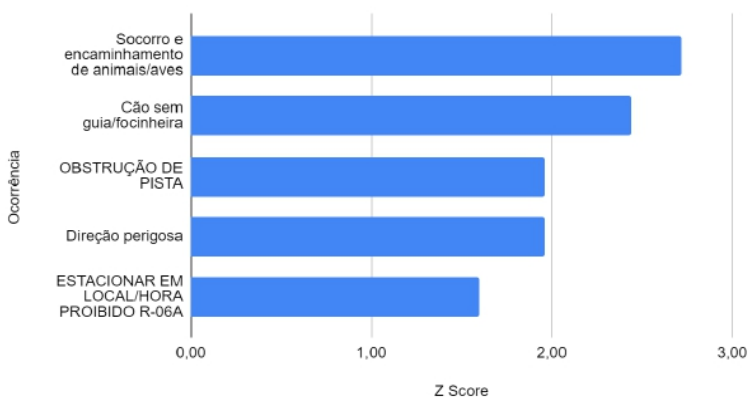


**Figura 10. Gráfico de Barras mostrando a quantidade de roubos em cada bairro.**

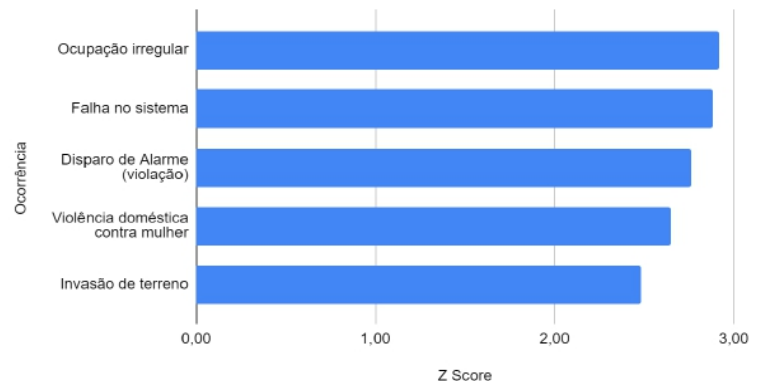
Em seguida foi contabilizado quantas ocorrências cada tipo teve em cada regional e normalizá-las para cada 10 mil habitantes que moram naquela regional. Com esse valor, calculou-se a porcentagem em que cada regional contribui para cada tipo de ocorrência e então calculou-se o desvio padrão dessa porcentagem entre as regionais.

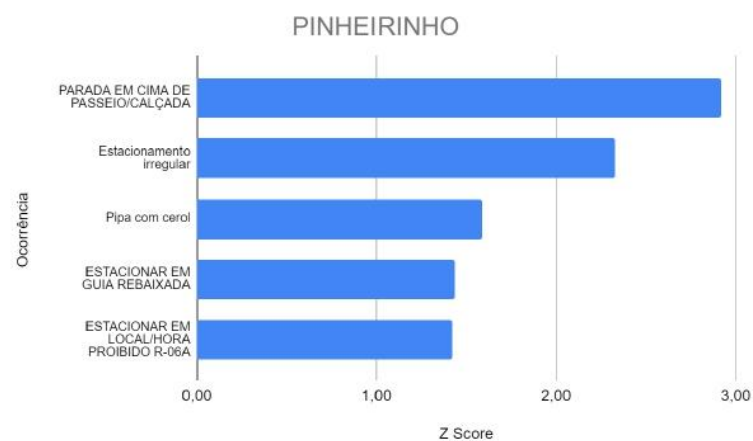
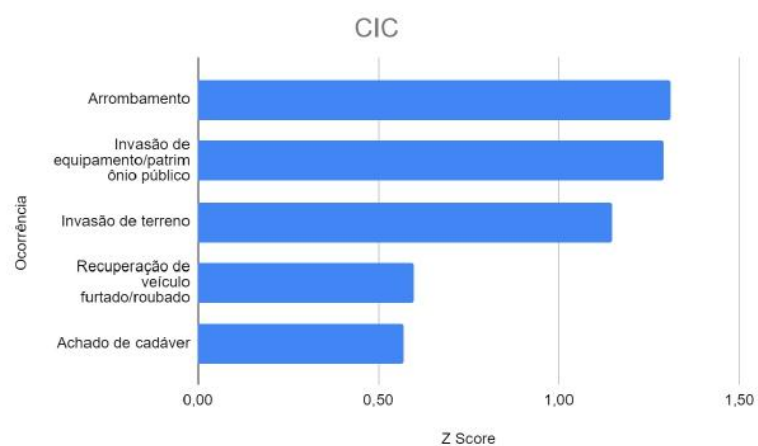
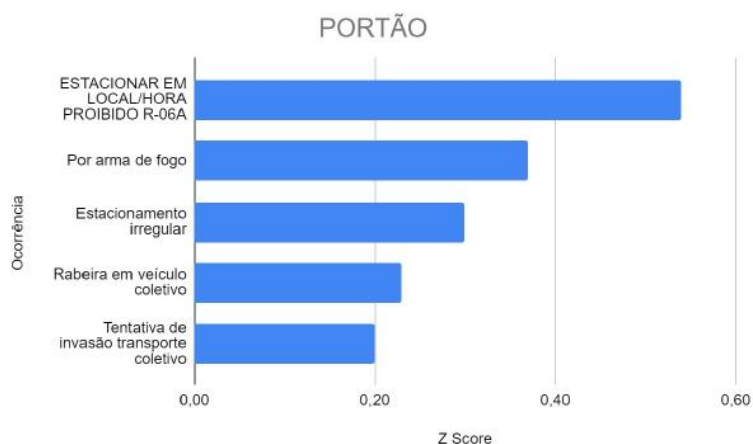
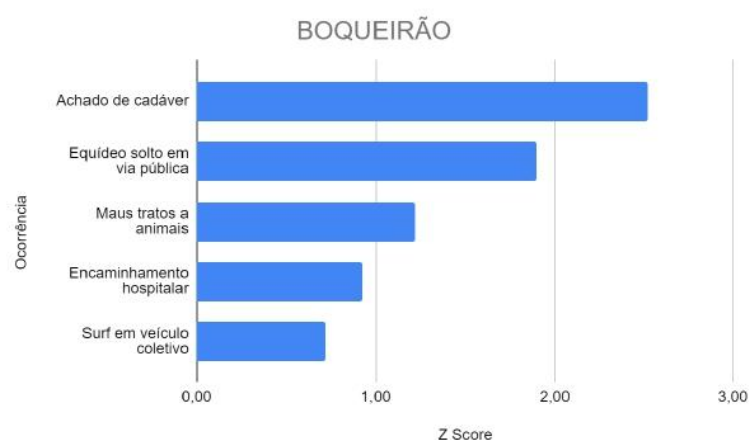
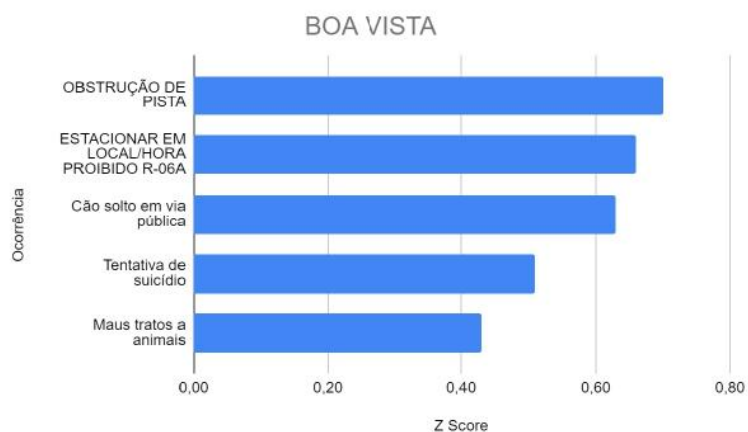
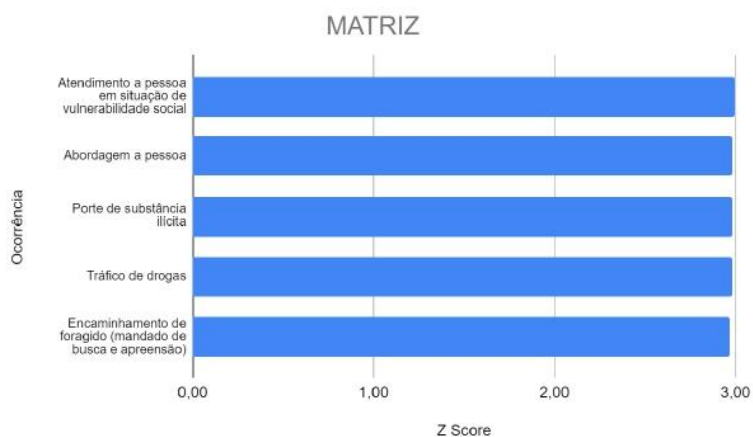
Em um caso onde os tipos de ocorrência são perfeitamente distribuídos, esperaria-se que cada regional contribuisse com 10% da quantidade normalizada de casos para cada 10 mil habitantes, mas este não foi o caso, sendo possível notar então, que em algumas regionais possuem certas particularidades que mais a diferenciam das demais, segue os gráficos dos Z-Scores obtidos nessa análise para cada regional:

**SANTA FELICIDADE**



**TATUQUARA**





Interessante notar como as regionais concentram tipos diferentes de ocorrências:

**Santa Felicidade:** Animais e direção imprópria de veículos.

**Tatuquara:** Terrenos, alarmes e violência doméstica contra mulheres.

**Matriz:** Drogas, atividades suspeitas e ajuda a moradores de rua.

**Boa vista:** Veículos, animais e tentativas de suicídio.

**Boqueirão:** Cadáver, animais, encaminhamento hospitalar e surf em ônibus.

**Portão:** Estacionamento de carros, porte de arma de fogo e relacionados aos ônibus.

**Cajuru:** Invasão de residência, animais e disparo de arma de fogo.

**Bairro Novo:** Arma de fogo, veículos roubados e invasão de patrimônio público.

**CIC:** Arrombamentos, invasões, recuperação de veículos furtados e cadáveres.

**Pinheirinho:** Estacionamento de carros e pipas com cerol.

## Perguntas de pesquisa e explorações iniciais

Existe uma relação entre a quantidade de crimes e os dias da semana? Com base no descrito na identificação de padrões iniciais é possível ver que não.

Existem horários mais propícios para os crimes? É possível ver que sim, já que existe uma concentração bem perceptível dos crimes no período noturno, e uma queda significativa no período da madrugada e manhã.

Qual o motivo das ocorrências de risco à vida crescerem bastante entre às 15h e 18h? Ainda não foi possível identificar.

Existe um horário propício para denúncias de crimes à mulher? Aparentemente não, apenas não ocorre muito no horário da madrugada.

É possível verificar uma relação entre população e renda e quantidade de crimes? Com base no que foi visto, é possível verificar que não há uma relação explícita, por não possuir um padrão específico.

## Discussão e próximos passos

Os dados de **crimes vs pessoa** reportados são parcialmente completos, foi possível entender que talvez não seja uma tarefa fácil relacionar os tipos de crimes com sexo, idade, etc. Porém vai ser possível criarmos mapas com as ocorrências e entender as relações com os bairros.

Também é possível verificar que é possível fazer uma grande quantidade de análises em relação aos tipos de crimes cometidos (qual o crime mais cometido em cada ajudbairro, se existe alguma relação entre os índices do bairro e o seu crime mais cometido, etc).

Verifica-se que também pode haver uma relação grande entre a quantidade proporcional de crimes em um determinado bairro, e sua distância em relação ao centro, isto é, verificando que os índices dos bairros não são muito afetados pela quantidade proporcional de crimes, pode ser que haja uma relação entre a distância.

Entendemos que não será necessário mudar o tema, foi adequado e essa análise exploratória nos deu muito insumo para vermos os pontos que podemos explorar. Será

interessante talvez buscar mais dados geográficos, dados da cidade de Curitiba (se possível).