

1. TEMA

O tema deste trabalho é o 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.

2. EQUIPE

Nome da equipe: Detetives de Dados.

Integrantes:

- Anderson Nogueira Silva, 2126516;
- Breno Moura de Abreu, 1561286;
- Giovana Viegas Barros, 2090066;
- Heitor Derder Trevisol, 1611810.

3. PERGUNTAS E HIPÓTESES DE PESQUISA

Nesta seção estão descritas as perguntas de pesquisa pensadas pelo grupo, bem como as hipóteses de pesquisa associadas a cada uma dessas perguntas.

- 3.1. Existe alguma relação entre a data, hora, mês, dia da semana e as ocorrências?
 - Nos fins de semana há mais ocorrências comparado com os demais dias da semana;
 - Nos meses de inverno (Junho-Setembro), há menos ocorrências em geral que nos demais meses do ano;
 - Há mais ocorrências no período noturno;
- 3.2. Podemos agrupar os bairros de Curitiba de acordo com os tipos de crimes mais comuns que ocorrem nestas regiões?

- Bairros mais afastados do centro possuem mais ocorrências violentas;
- Área territorial, população, renda per capita influenciam no número de ocorrências.

3.3. Qual é a tendência para o futuro em relação à criminalidade em Curitiba? Existem tipos de crimes com crescimento ou redução em números de ocorrências?

- A cidade está ficando mais violenta com o passar dos anos;
- Com o aumento da população há aumento de ocorrências;

3.4. Existiu em algum momento tipos de ocorrências que apresentaram um comportamento anômalo?

- Houve uma queda acentuada de ocorrências criminais no ano de 2020 por conta da pandemia de COVID-19;
- As mudanças da política vigente interferem no tipo e número de ocorrências registradas.

4. DADOS E MODELOS

Para montar os modelos, responder as perguntas e averiguar as hipóteses a fonte primária de dados que estaremos utilizando será, como discutido em documentos anteriores, a base de dados SiGesGuarda [\[1\]](#), conjunto este relativo aos registros de ocorrências por parte da Guarda Municipal de Curitiba no período de 2009 até 2022.

Como complemento também utilizaremos dados de indicadores para as questões de normalização dos dados, como por exemplo população da cidade (inteira ou por bairro), área (em m^2) dos bairros, renda per capita média, preço médio do m^2 . Para os valores referentes a indicadores (renda, população por bairro) estaremos usando os dados compilados no documento: ESTUDO TEMÁTICO 2 [\[2\]](#). Para os dados de área dos bairros e população da cidade usamos a população estimada da cidade de Curitiba pelo IBGE [\[3\]](#), e por fim para o valor do m^2 utilizaremos a estimativa montada pela empresa Agente Imóvel [\[4\]](#).

Para cada hipótese estabelecida, será definido um modelo de análise para responder às perguntas de pesquisa, conforme apresentado na Tabela 1.

Para analisar as relações entre as ocorrências e o tempo (hora, dia e mês), serão plotados diversos gráficos com diferentes níveis de granularidade temporal. A análise dos gráficos e das médias comparadas irá permitir verificar em quais períodos de tempo as ocorrências estão mais concentradas. Isso permite responder à pergunta 3.1 e verificar se suas hipóteses são verdadeiras.

Para a pergunta 3.2, será utilizado um modelo de aprendizado de máquina para a clusterização dos bairros. O modelo K-means permitirá agrupar os bairros de forma não supervisionada, e a criação de um dendrograma possibilitará determinar qual é o melhor número de agrupamentos para a análise. A partir dos grupos formados, uma análise mais aprofundada será realizada para verificar as hipóteses. Além disso, para verificar a influência da área territorial, renda per capita e população dos bairros na criminalidade será feito uso de regressão logística.

Em relação ao estudo de tendências para a pergunta 3.3, será utilizado técnicas estatísticas como médias móveis assim como regressão para determinar as tendências em relação à criminalidade ao longo do tempo. Gráficos serão criados para permitir a visualização dos resultados encontrados na utilização das técnicas citadas.

Para a detecção de comportamentos anômalos, apontados na pergunta 3.4, será feito o uso de modelos de detecção de *outliers* e a análise das médias para verificar períodos de tempo onde houve alteração nos índices de criminalidade.

PERGUNTAS DE PESQUISA	HIPÓTESES	MODELO
Existe alguma relação entre a data, hora, mês, dia da semana e as ocorrências?	Nos fins de semana há mais ocorrências comparado com os demais dias da semana;	Médias; análise dos gráficos.
	Nos meses de inverno (Junho-Setembro), há menos ocorrências em geral que nos demais meses do ano;	Médias; análise dos gráficos.
	Há mais ocorrências no período noturno;	Médias; análise dos gráficos.
Podemos agrupar os bairros de Curitiba de acordo com os tipos de crimes mais comuns que ocorrem nestas regiões?	Bairros mais afastados do centro possuem mais ocorrências violentas;	Agrupamento.
	Área territorial, população, renda per capita influenciam no número de ocorrências.	Regressão Logística.
Qual é a tendência para o futuro em relação à criminalidade em Curitiba? Existem tipos de crimes com crescimento ou redução em números de ocorrências?	A cidade está ficando mais violenta com o passar dos anos;	Regressão; médias móveis.
	Com o aumento da população há aumento de ocorrências;	Regressão.
Existiu em algum momento tipos de ocorrências que apresentaram um comportamento anômalo?	Houve uma queda acentuada de ocorrências criminais no ano de 2020 por conta da pandemia de COVID-19;	Deteção de outliers.
	As mudanças da política vigente interferem no tipo e número de ocorrências registradas.	Médias;

Tabela 1: Modelos utilizados para análise das hipóteses.

5. CRONOGRAMA

- 5.1. Inicialmente, por meio de gráficos construídos na etapa da Análise exploratória e de novas visualizações relevantes, inferir se as hipóteses estabelecidas são plausíveis;
- 5.2. Aplicar modelos de aprendizado de máquina para as hipóteses definidas sobre o agrupamento dos bairros de Curitiba pelo tipo de ocorrência;
- 5.3. Por meio de técnicas estatísticas validar as hipóteses sobre a tendência da criminalidade na cidade de Curitiba;
- 5.4. Construir modelos de regressão para verificar o efeito que mudanças/eventos causam no comportamento das ocorrências registradas na cidade de Curitiba;
- 5.5. Estabelecer resultados prévios das hipóteses definidas para apresentação de *checkpoint*;
- 5.6. Aprofundar e refinar as análises como um todo para as hipóteses mais relevantes;
- 5.7. Elaboração da apresentação com os resultados finais considerados pela equipe;
- 5.8. Elaboração do relatório contendo o resultado da condução do trabalho.

6. REFERÊNCIAS

[1] 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>>.

[2] Observatório do Trabalho de Curitiba. **ESTUDO TEMÁTICO 2**: Perfil demográfico e socioeconômico dos bairros agregados de Curitiba. Fevereiro 2016. Disponível em: <https://www.coreconpr.gov.br/wp-content/uploads/2016/07/bairros.pdf>. Acesso em: 10 maio 2023.

[3] IBGE. **Panorama Cidade de Curitiba, Paraná**. Disponível em: <<https://cidades.ibge.gov.br/brasil/pr/curitiba/panorama>>.

[4] Agente Imóvel. **Valor do m2 Curitiba, PR**. Disponível em: <<https://www.agenteimovel.com.br/mercado-imobiliario/a-venda/pr/curitiba/#>>.