

Exploração e Visualização de Padrões Criminais

Dirleide Silva¹, Priscylla Silva^{1,2}, Leonardo B. M. Silva¹

¹Instituto Federal de Alagoas - Campus Rio Largo

²Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo (USP) – São Carlos, SP – Brasil

gas27@aluno.ifal.edu.br, priscylla.silva@usp.br

leonardo.silva@ifal.edu.br

Abstract. *Criminality is one of the problems that most affect our society. In recent years, data science techniques have been used to help mitigate this problem. It is already known in the literature that crimes have a temporal and spatial pattern in cities. However, most works focus on analyzing data from a single city. This paper presents an investigation based on visual analysis to compare criminal patterns in Chicago and San Francisco. The results show that their criminal patterns differ.*

Resumo. *A criminalidade é um dos problemas que mais afetam a nossa sociedade. Nos últimos anos, técnicas de ciência de dados têm sido utilizadas para ajudar a mitigar esse problema. Já é conhecido na literatura sobre o tema, que os crimes apresentam um padrão temporal e espacial nas cidades. Entretanto, a maioria dos trabalhos estão concentrados na análise de dados de uma única cidade. Este trabalho apresenta uma investigação baseada em análise visual para comparar padrões criminais em duas cidades, Chicago e São Francisco. Os resultados encontrados mostram que seus padrões criminais são distintos.*

1. Introdução

A ocorrência de crimes no contexto urbano é um problema mundial [Nadai et al. 2020]. Recursos tecnológicos podem ser aplicados com o intuito de buscar soluções para ajudar a mitigar esse problema. Em particular, a tecnologia pode ser empregada na análise automática da massiva quantidade de dados disponíveis na área de segurança pública, tais como, registros de ocorrências criminais e informações espaço temporais referentes às localidades onde as ocorrências aconteceram [Han et al. 2020]. Este grande volume de dados pode ser utilizado para a geração de conhecimento e auxiliar na tomada de decisões inteligentes. Cidades americanas como Chicago e São Francisco disponibilizam de forma aberta os dados de ocorrências criminais. Com a disponibilidade dos dados, é possível utilizar técnicas de ciência de dados para análise e descoberta de padrões criminais. Diversos trabalhos na literatura focam no uso de análise visual para descoberta desses padrões, entretanto, eles apresentam uma série de limitações, como, por exemplo, serem restritos a uma única cidade.

Este trabalho tem como objetivo realizar uma análise comparativa entre padrões criminais de duas cidades americanas, São Francisco e Chicago. Por meio de técnicas de análise visual é possível investigar se as cidades possuem padrões temporais e espaciais semelhantes ou distintos.

2. Materiais e Métodos

Os dados utilizados neste estudo são base de dados reais de duas cidades americanas, Chicago e São Francisco. Ambas as bases foram coletadas dos sites oficiais das prefeituras [Chicago-Data-Portal 2023, DataSF 2023]. A base de dados de Chicago possuía 7.749.627 registros de janeiro de 2001 a fevereiro de 2023. A base continha 22 atributos relacionados à localização geográfica, horário do crime, códigos criminais americanos e características da ação criminal (por exemplo, se o criminoso foi preso). Na etapa de limpeza e pré-processamento, foram mantidos apenas os dados de localização e horário, além da categoria do crime. Também foram removidos os dados faltantes, duplicados e com latitude ou longitude fora dos limites da cidade.

A etapa de limpeza e pré-processamento na base da cidade de São Francisco foi um pouco mais complexa, pois ela é disponibilizada dividida em dois arquivos, um contendo os dados de 2003 até maio de 2018 e outro de janeiro de 2018 até a data atual (neste estudo, essa base termina em fevereiro de 2023). Porém, os dois arquivos possuem quantidade de atributos diferentes, assim como também possuem nomes diferentes. A base de 2003 a 2018 possuía 1.882.474 registros e 35 atributos, enquanto a base de 2018 a 2023 possuía 703.772 registros e 34 atributos. As duas bases de dados foram integradas e os conflitos de atributos resolvidos. As mesmas etapas de limpeza e pré-processamento descritas para a cidade de Chicago também foram aplicadas para São Francisco.

Devido ao funcionamento da legislação americana, onde cada estado possui sua própria categorização de crimes, optou-se por selecionar apenas as categorias que estivessem entre as 10 mais frequentes em ambas as bases e também tivessem a mesma definição em ambas as cidades. Esses critérios são importantes, pois diferentes tipos de crimes podem possuir padrões distintos, por tanto, comparar tipos diferentes ou o conjunto completo de todos os tipos de crimes pode levar a conclusões errôneas. Apenas a categoria *assault* atendia aos dois critérios, sendo este tipo de crime definido como “ameaça física de violência contra outro indivíduo”.

Como as duas bases de dados possuem intervalos de tempo diferentes, optamos por trabalhar com os registros dos anos de 2008 a 2015. A restrição a 2015 ocorreu porque a base de São Francisco não possui registros da categoria *assault* após esse ano. Ao final a base de Chicago continha 161.022 registros e 9 atributos (categoria do crime, data, ano, mês, dia, horário, dia da semana, latitude, longitude). Já a base da cidade de São Francisco continha 73.006 registros e 9 atributos (os mesmo que Chicago). Todas as análises foram realizadas utilizando a linguagem de programação Python com uso das bibliotecas: Numpy, Pandas, GeoPandas, Matplotlib, Seaborn e Folium.

3. Resultados e Discussões

Nesta seção serão apresentados os resultados obtidos da análise exploratória dos dados e da comparação visual entre os padrões de crimes das duas cidades. Devido a limitações de espaço, apenas os resultados mais relevantes serão reportados.

Como apresentado na Figura 1, a tendência de ocorrências de crimes nas duas cidades são distintas. A cidade de Chicago apresenta uma queda anual no número de registros para o período analisado, já a cidade de São Francisco vinha apresentando uma queda até o ano de 2012, quando o número de crimes inicia uma tendência de aumento.

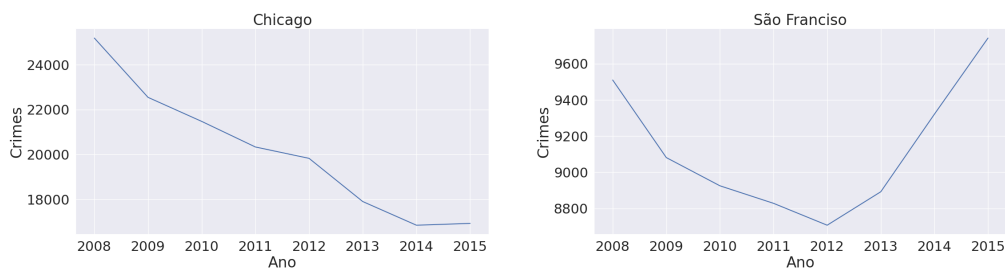


Figura 1. Quantitativo de crimes de *assault* por ano.

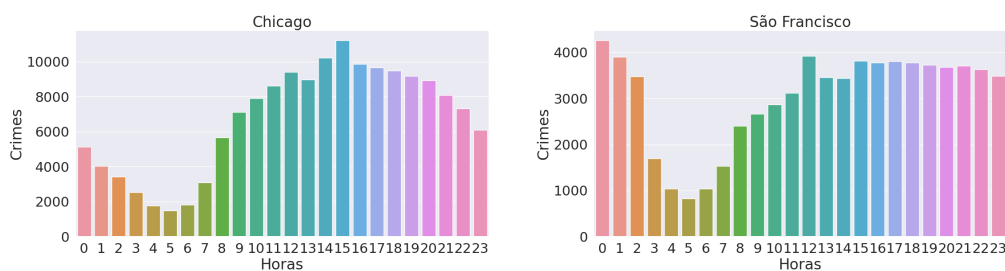


Figura 2. Quantitativo de crimes de *assault* ao longo das horas do dia.

Com relação a ocorrência do crime ao longo do dia, a Figura 2 mostra um comparativo entre as duas cidades. Pode-se observar que ambas as cidades apresentam uma redução no número de ocorrências entre as duas e as sete horas da manhã. Entretanto, São Francisco possui uma ocorrência de crime mais constante entre o fim da tarde e ao longo da noite (das 15 às 23 horas). Já a cidade de Chicago, para o mesmo período, apresenta uma redução gradativa do número de ocorrências.

No que diz respeito à distribuição de ocorrências durante a semana, as cidades apresentam comportamentos opostos (Figura 3). Em Chicago, a maioria dos crimes de *assault* ocorrem durante os dias úteis da semana (de segunda a sexta), enquanto em São Francisco a maioria das ocorrências ocorrem no final de semana.

Também é possível constatar que a distribuição espacial do crime possui padrões distintos. Na cidade de São Francisco, a ocorrência de *assault* se concentra na região nordeste da cidade (Figura 4), enquanto em Chicago as ocorrências estão espalhadas por toda a cidade (Figura 5). A região de São Francisco onde esse tipo de crime se concentra é também a região com maior pontos turísticos.

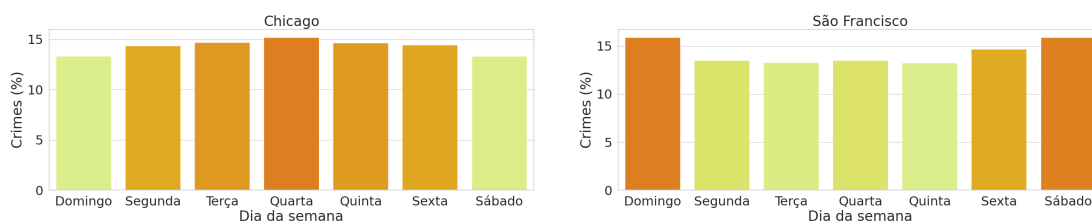


Figura 3. Quantitativo de crimes de *assault* por dia da semana.

