



Universidade Tecnológica Federal do Paraná
Campus Campo Mourão
Departamento de Computação - DACOM
Prof. Dr. Eduardo Pena
Disciplina: Ciência de Dados



Projeto Final: Análise de Dados Aplicada

Etapa 1 – Definição do Problema

Aluno: Elder Henrique Alves Correia
RA: 2222698

Novembro / 2025

1. Identificação do Grupo e Repositório Git

- Integrantes: Elder Henrique Alves Correia
- Repositório Git:
https://github.com/ElderCorreia7/projeto_final_ciencia_de_dados.git

2. Descrição do Dataset

- **Título:** NYPD Arrests Data (*Historic*).
- **Fonte:** [NYC Open Data](#).
- **Categoria:** Segurança Pública.
- **Tamanho:** 5,99 milhões de registros e 19 preditores (categóricas, numéricas, geográficas, datas).
- **Estrutura e conteúdo:** Cada registro corresponde a uma prisão realizada pelo NYPD em Nova York, incluindo o tipo de crime, localização, hora da prisão, informações do suspeito, entre outros aspectos.
- **Formato utilizado no projeto:** CSV.
- **Descrição:** O *dataset* selecionado é o NYPD Arrest Data (*Historic*), disponibilizado pelo portal NYC Open Data, mantido pela prefeitura de Nova York. Ele contém registros detalhados de todas as prisões realizadas pelo Departamento de Polícia de Nova York (NYPD) desde 2006, incluindo tipo de crime, localização, data/hora, dados demográficos dos suspeitos, entre outros.

3. Contexto e Justificativa

A criminalidade urbana é um dos temas mais relevantes dentro do campo de políticas públicas, segurança e análise social. Grandes cidades como Nova York registram milhares de prisões todos os anos, envolvendo pessoas de diferentes perfis e tipos de delitos.

O dataset do NYPD, contém registros detalhados de prisões realizadas em Nova York desde 2006, incluindo informações sobre tipo de crime, localização, data/hora e características dos suspeitos. Com base nesses dados, a análise a ser realizada neste trabalho, buscará investigar a relação entre o tipo de crime registrado (OFNS_DESC) e as características demográficas dos suspeitos, como sexo (PERP_SEX) e faixa etária (AGE_GROUP). Deste modo, serão utilizadas análises exploratórias, tabelas cruzadas, visualizações e modelagem preditiva de classificação para identificar padrões de comportamento e tendências criminais específicas de cada grupo demográfico.

A análise proposta tem impacto prático, pois:

- A análise ajuda a entender como diferentes faixas etárias se relacionam com tipos específicos de crimes;
- Permite identificar como homens e mulheres se envolvem em diferentes tipos de crimes;
- Fornece um exemplo de como trabalhar com grandes bases de dados;

4. Perguntas de Pesquisa

O objetivo desta pesquisa é investigar se existe relação entre o tipo de crime registrado (OFNS_DESC) e características demográficas dos suspeitos, como sexo (PERP_SEX) e faixa etária (AGE_GROUP).

Para isso, busca-se identificar padrões de criminalidade associados a diferentes perfis demográficos, por meio de análises estatísticas, tabelas cruzadas, visualizações exploratórias e modelos preditivos. Espera-se que os resultados ofereçam uma compreensão mais clara sobre como esses grupos se relacionam com determinados tipos de crimes, revelando tendências e padrões que muitas vezes passam despercebidos em análises tradicionais.

5. Hipóteses Testáveis

A seguir temos as seguintes hipóteses:

H1 — O tipo de crime está associado ao sexo do suspeito

- Espera-se observar diferenças significativas na distribuição de crimes entre suspeitos masculinos e femininos, indicando que certos tipos de crimes são mais frequentes em um grupo específico.

H2 — O tipo de crime está associado à faixa etária do suspeito

- Supõe-se que determinadas faixas etárias estejam mais relacionadas a crimes específicos, refletindo padrões de comportamento e dinâmica criminal ao longo da vida.