

**Universidade Presbiteriana Mackenzie**  
**Tecnologia em Ciência de Dados - Projeto Aplicado 2**

Franciele do Nascimento  
Leandro Rodrigues Dos Santos  
Luiz Eduardo de Mei Salvador Coelho  
Matheus Neves de Castro

**Projeto Aplicado II**

SÃO PAULO  
1º Semestre de 2024

Autores:

Leandro Rodrigues Dos Santos / 23019689 / lers.138@gmail.com

Franciele do Nascimento / 10414598 / f.paterni@hotmail.com

Luiz Eduardo de Mei Salvador Coelho / 23024585 / Luiz02coelho@gmail.com

Matheus Neves de Castro / 10415190 / 10415190@mackenzista.com.br

## Projeto Aplicado II

Professor: Anderson Adaime de Borba

SÃO PAULO

1º Semestre de 2024

## Índice

Introdução .....	4
Objetivos e Metas .....	5
DataSet e Metadados .....	6
Repositório .....	6
Cronograma geral .....	7
Bibliotecas usadas no projeto .....	8
Definir a base de dados e a sua análise exploratória .....	8
Tratar a base de dados (preparação e treinamento).....	9
Definir e descrever as bases teóricas dos métodos analíticos.....	9
Definir e descrever como será calculada a acurácia.....	10

## 1. Introdução

A cidade de São Paulo é frequentemente citada como uma das mais perigosas do país. Milhões de habitantes se locomovem na agitada e imparável capital paulista. Ser paulistano é saber que você não pode andar na rua com o celular na mão, não pode sair a noite no centro e nem tirar uma selfie em plena luz do dia, em determinadas circunstâncias.

Quando falamos de uma das maiores cidades do mundo, qualquer desafio se torna muito mais complexo. Dando um zoom na segurança pública, temos diversos fatores de complexidade que vão desde a escassez de recursos por parte da união até o desafio de alocar de forma inteligente a força policial para minimizar o número de incidentes.

De acordo com os dados da SSP (secretaria de segurança pública) de São Paulo, a maioria esmagadora dos crimes que ocorrem na capital é feita em via pública. Temos ainda os dados de que estão a disposição da capital 40 mil policiais militares e cerca de 4 mil viaturas.

Dentro desse cenário, proporemos uma abordagem de otimização dos recursos atualmente disponíveis visando a maximização da segurança com o mínimo de recursos possíveis alocados. Essa otimização se baseará nos dados de BO.s da cidade e usará técnicas de análise de dados e de ML avançadas para entender os comportamentos padrões e melhorar a alocação dos recursos.

## 2. Objetivos e Metas

O trabalho busca:

Melhor compreensão do estilo do crime na cidade de São Paulo com análises descritivas:

- Como o crime tem evoluído ao longo do tempo?
- Quais os horários mais perigosos na cidade?
- Quais os tipos de crimes mais recorrentes?
- Quais os locais com mais incidência de crimes?
- Quais os bairros mais perigosos?

Construir uma sólida base de modelagem e um modelo acurado para prever a melhor alocação de recursos da cidade:

- Feature Engineering de qualidade com as variáveis corretas para a modelagem.
- Testar diferentes abordagens de modelagem para melhorar a acurácia com Reinforcement Learning (MARL).
- Avaliar o desempenho do modelo em termos de acurácia, confiabilidade e interpretabilidade.
- Selecionar o modelo que melhor se adapta ao problema com melhores resultados

Uma vez modelado, ter um output claro de:

- Determinar o número mínimo de viaturas necessário para garantir a segurança da cidade, com base nos resultados da modelagem.
- Identificar o ponto de inflexão, onde adicionar mais viaturas não resulta em uma redução significativa da criminalidade.
- Quantificar de acordo com os dados qual a redução possível da taxa de criminalidade aplicando a inteligência nos dados.

### 3. DataSet e Metadados

Referências de aquisição do dataset:

Foi adquirido através do site: <https://www.ssp.sp.gov.br/estatistica>

Descrição da origem:

Site oficial da secretaria de segurança pública do estado de São Paulo

Descrição do dataset:

Conjunto completo de todos os BOs do estado de São Paulo com o tipo do crime, hora, logradouro, latitude e longitude, etc.

Foto do dataset:

NOME DEPARTAMENTO	NOME SECCIONAL	NOME DELEGACIA	CIDADE	NUM_BO	ANO_BO	DATA_COMUNICACAO_BO	DATA_OCORRENCIA_BO	HORA_OCORRENCIA_BO	DESCR_PERIODO	DESCR_TIPO
DECAP	DELSEC-8 SÃO MATEUS	084 DOM SÃO MATEUS	S.PAULO	AC0851	2022	03/09/2022	03/09/2022	15:30:00	Residência	Via pública
DECAP	DELSEC-8 SÃO MATEUS	544 D.P. CIO. TRIGANTES	S.PAULO	7	2022	03/01/2022	31/12/2021	15:30:00	EM HORA INCERTA	Terminal/Es
DECAP	DELSEC-14 LESTE	299 D.P. VILA PRUDENTE	S.PAULO	8	2022	04/01/2022	30/12/2021	15:30:00	Terminal/Es	Terminal/Es
DECAP	DELSEC-14 CENTRO	019 D.P. SE	S.PAULO	8	2022	03/01/2022	03/01/2022	15:30:00	Via pública	Via pública
DECAP	DELSEC-14 CENTRO	069 D.P. CAMBUÍ	S.PAULO	8	2022	03/01/2022	03/01/2022	15:30:00	Via pública	Via pública
DECAP	DELSEC-21 SUL	170 D.P. DOUTOR ALDO GALIANO	S.PAULO	11	2022	04/01/2022	04/01/2022	11:40:00	Via pública	Via pública
DECAP	DELSEC-71 ITAQUERA	689 D.P. LAGADOU	S.PAULO	13	2022	03/01/2022	31/12/2021	11:45:00	Comércio e	Comércio e
DECAP	DELSEC-14 CENTRO	089 D.P. BRAS	S.PAULO	12	2022	02/01/2022	02/01/2022	01:20:00	Via pública	Via pública
DOPE-DEPTO DP POL. ESTRAT.	DIV.POL.PORTO/AERO/PROT.TURISTA	DOPE - 014 DP DEATUR CAPITAL	S.PAULO	12	2022	09/01/2022	08/01/2022	17:00:00	Comércio e	Comércio e
DECAP	DELSEC-14 CENTRO	124 D.P. PARÍ	S.PAULO	13	2022	03/01/2022	25/12/2021	17:00:00	Terminal/Es	Terminal/Es
DECAP	DELSEC-TARTININGA	DEL.POL.SÃO MIGUEL ARCANJO	S.PAULO	14	2022	04/01/2022	28/12/2021	21:23:00	Via pública	Via pública
DOPE-DEPTO DP POL. ESTRAT.	DIV.POL.PORTO/AERO/PROT.TURISTA	DOPE - 014 DP DEATUR CAPITAL	S.PAULO	14	2022	10/01/2022	10/01/2022	19:11:00	EM HORA INCERTA	Terminal/Es
DECAP	DELSEC-14 CENTRO	774 D.P. SANTA CECÍLIA	S.PAULO	14	2022	03/01/2022	03/01/2022	12:05:00	Via pública	Via pública
DECAP	DECAP - SEDE	DEL.POL.METROPOLITANO	S.PAULO	15	2022	04/01/2022	04/01/2022	07:00:00	Terminal/Es	Terminal/Es
DECAP	DELSEC-14 CENTRO	789 D.P. JARDINS	S.PAULO	17	2022	02/01/2022	02/01/2022	04:50:00	Residência	Residência
DECAP	DELSEC-14 CENTRO	019 D.P. SE	S.PAULO	21	2022	03/01/2022	03/01/2022	08:05:00	Via pública	Via pública
DECAP	DELSEC-21 SUL	170 D.P. DOUTOR ALDO GALIANO	S.PAULO	24	2022	06/01/2022	05/01/2022	20:20:00	Via pública	Via pública
DECAP	DELSEC-14 CENTRO	019 D.P. SE	S.PAULO	26	2022	05/01/2022	21/12/2021	21:00:00	Condomíni	Condomíni
DECAP	DELSEC-14 CENTRO	059 D.P. ACIMACAO	S.PAULO	28	2022	06/01/2022	05/01/2022	20:40:00	Via pública	Via pública
DECAP	DELSEC-14 CENTRO	089 D.P. BRAS	S.PAULO	30	2022	03/01/2022	03/01/2022	16:50:00	Comércio e	Comércio e
DECAP	DELSEC-14 CENTRO	019 D.P. SE	S.PAULO	30	2022	03/01/2022	03/01/2022	18:00:00	Via pública	Via pública
DECAP	DELSEC-14 CENTRO	019 D.P. SE	S.PAULO	31	2022	03/01/2022	03/01/2022	18:50:00	Via pública	Via pública
DECAP	DELSEC-14 CENTRO	999 D.P. CAMPO GRANDE	S.PAULO	31	2022	03/01/2022	03/01/2022	17:00:00	Via pública	Via pública
DECAP	DELSEC-14 CENTRO	019 D.P. SE	S.PAULO	33	2022	06/01/2022	24/12/2021	16:30:00	Via pública	Via pública
DECAP	DECAP - SEDE	DEL.POL.METROPOLITANO	S.PAULO	34	2022	10/01/2022	08/01/2022	07:40:00	Terminal/Es	Terminal/Es
DOPE-DEPTO DP POL. ESTRAT.	DIV.POL.PORTO/AERO/PROT.TURISTA	DOPE - 014 DP DEATUR CAPITAL	S.PAULO	34	2022	17/01/2022	17/01/2022	10:53:00	Comércio e	Comércio e
DECAP	DELSEC-14 CENTRO	059 D.P. ACIMACAO	S.PAULO	36	2022	07/01/2022	06/01/2022	12:30:00	Via pública	Via pública
DECAP	DELSEC-71 ITAQUERA	679 D.P. JARDIM ROBU	S.PAULO	38	2022	10/01/2022	07/01/2022	16:00:00	Via pública	Via pública
DECAP	DELSEC-14 CENTRO	059 D.P. ACIMACAO	S.PAULO	37	2022	07/01/2022	06/01/2022	11:30:00	Via pública	Via pública
DIHOL - DEPTO DE INTELIGENCIA	DELEGACIA ELETRONICA	DELEGACIA ELETRONICA	S.PAULO	37	2022	01/01/2022	31/12/2021	11:30:00	Via pública	Via pública
DECAP	DELSEC-71 ITAQUERA	644 D.P. CID AE CARVALHO	S.PAULO	38	2022	05/01/2022	05/01/2022	12:55:00	Via pública	Via pública
DECAP	DELSEC-14 CENTRO	089 D.P. BRAS	S.PAULO	38	2022	04/01/2022	04/01/2022	08:40:00	Via pública	Via pública
DECAP	DELSEC-14 CENTRO	019 D.P. SE	S.PAULO	39	2022	06/01/2022	05/01/2022	10:40:00	Via pública	Via pública
DECAP	DELSEC-14 CENTRO	019 D.P. SE	S.PAULO	44	2022	06/01/2022	05/01/2022	12:30:00	Escritório	Escritório
DOPE-DEPTO DP POL. ESTRAT.	DIV.POL.PORTO/AERO/PROT.TURISTA	DOPE - 014 DP DEATUR CAPITAL	S.PAULO	46	2022	21/01/2022	21/01/2022	14:20:00	Via pública	Via pública
DECAP	DECAP - SEDE	DEL.POL.METROPOLITANO	S.PAULO	47	2022	13/01/2022	13/01/2022	05:25:00	Terminal/Es	Terminal/Es
DOPE-DEPTO DP POL. ESTRAT.	DIV.POL.PORTO/AERO/PROT.TURISTA	DOPE - 014 DP DEATUR CAPITAL	S.PAULO	47	2022	06/01/2022	06/01/2022	05:25:00	Escritório	Escritório
DECAP	DELSEC-14 CENTRO	019 D.P. SE	S.PAULO	47	2022	06/01/2022	06/01/2022	16:30:00	Via pública	Via pública
DECAP	DELSEC-14 CENTRO	369 D.P. VILA MARIANA	S.PAULO	50	2022	12/01/2022	12/01/2022	16:00:00	Terminal/Es	Terminal/Es
DECAP	DELSEC-8 SÃO MATEUS	669 D.P. JONIA BRIGANDIUA	S.PAULO	50	2022	10/01/2022	08/01/2022	20:30:00	Terminal/Es	Terminal/Es
DOPE-DEPTO DP POL. ESTRAT.	DIV.POL.PORTO/AERO/PROT.TURISTA	DOPE - 014 DP DEATUR CAPITAL	S.PAULO	50	2022	22/01/2022	21/01/2022	17:30:00	Via pública	Via pública
DECAP	DELSEC-14 CENTRO	089 D.P. BRAS	S.PAULO	51	2022	05/01/2022	22/12/2021	13:50:00	Condomíni	Condomíni
DECAP	DECAP - SEDE	DEL.POL.METROPOLITANO	S.PAULO	52	2022	14/01/2022	14/01/2022	09:45:00	Terminal/Es	Terminal/Es
DECAP	DELSEC-21 SUL	839 D.P. PARQUE BRISTOL	S.PAULO	52	2022	18/01/2022	18/01/2022	13:40:00	Via pública	Via pública
DECAP	DELSEC-21 SUL	579 D.P. AMERIGANDOLIS	S.PAULO	52	2022	18/01/2022	18/01/2022	11:00:00	Via pública	Via pública
DOPE-DEPTO DP POL. ESTRAT.	DIV.POL.PORTO/AERO/PROT.TURISTA	DOPE - 014 DP DEATUR CAPITAL	S.PAULO	56	2022	24/01/2022	24/01/2022	14:00:00	Via pública	Via pública
DECAP	DECAP - SEDE	DEL.POL.METROPOLITANO	S.PAULO	59	2022	17/01/2022	15/01/2022	11:45:00	Terminal/Es	Terminal/Es
DECAP	DELSEC-14 CENTRO	089 D.P. BRAS	S.PAULO	59	2022	05/01/2022	05/01/2022	09:00:00	Via pública	Via pública
DECAP	DELSEC-14 CENTRO	019 D.P. SE	S.PAULO	60	2022	07/01/2022	07/01/2022	10:20:00	Via pública	Via pública
DOPE-DEPTO DP POL. ESTRAT.	DIV.POL.PORTO/AERO/PROT.TURISTA	DOPE - 014 DP DEATUR CAPITAL	S.PAULO	60	2022	25/01/2022	25/01/2022	16:54:00	Comércio e	Comércio e
DECAP	DELSEC-14 CENTRO	019 D.P. SE	S.PAULO	71	2022	08/01/2022	08/01/2022	15:20:00	Via pública	Via pública
DOPE-DEPTO DP POL. ESTRAT.	DIV.POL.PORTO/AERO/PROT.TURISTA	DOPE - 014 DP DEATUR CAPITAL	S.PAULO	71	2022	29/01/2022	28/01/2022	07:00:00	Via pública	Via pública

### 4. Repositório

O link para o repositório do trabalho é: <https://github.com/MatNew/Projeto->

[Aplicado-2](#)

## 5. Cronograma Geral

Cronograma Geral do Projeto Aplicado II				
	Tarefa	Responsável	Status	Concluído em:
Etapa 1 : Kick-Off	Definir o grupo de trabalho.	TODOS	CONCLUIDO	19/02/24
	Definir as premissas do projeto: definição da empresa, área de atuação e apresentação dos dados que serão utilizados (imagem ou texto).	TODOS	CONCLUIDO	19/02/24
	Determinar objetivos e metas.	TODOS	CONCLUIDO	19/02/24
	Criar um cronograma de atividades.	LEANDRO	CONCLUIDO	19/02/24
	Criação do GITHUB	MATHEUS	CONCLUIDO	01/03/24
ETAPA 2 - *Definição do produto*	Definir quais bibliotecas (pacotes) da linguagem python e qual repositório no GitHub devem ser usados para iniciar a execução colaborativa do trabalho.	LUIZ	Não iniciado	
	Definir a base de dados e a sua análise exploratória.	LUIZ	Não iniciado	
	Tratar a base de dados (preparação e treinamento).	LUIZ	Não iniciado	
	Definir e descrever as bases teóricas dos métodos analíticos.	LUIZ	Não iniciado	
	Definir e descrever como será calculada a acurácia.	LUIZ	Não iniciado	
ETAPA 3 - *Storytelling*	Consolidar os resultados do método analítico definido na etapa anterior, aplicado à base de dados definida com padrão.	MATHEUS	Não iniciado	
	Aplicar as medidas de acurácia para verificar o desempenho dos métodos definidos na etapa anterior.	MATHEUS	Não iniciado	
	Descrever os resultados preliminares, apresentando um produto gerado, e descrevendo um possível modelo de negócios.	LEANDRO E FRANCIELE	Não iniciado	

	Esboçar o Storytelling	LEANDRO E FRANCIELE	Não iniciado	
ETAPA 4 - *Encerramento*	Relatório técnico do projeto.	LEANDRO	Não iniciado	
	Apresentação do Storytelling Pode ser um arquivo PPT ou similares	LEANDRO	Não iniciado	
	Repositório no GitHub com todos arquivos do projeto	MATHEUS	Não iniciado	
	Vídeo com a apresentação no YouTube, usando a técnica do Storytelling	FRANCIELE	Não iniciado	

## 6. Bibliotecas usadas no projeto

Para o trabalho, usaremos as seguintes bibliotecas python:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import re
import unicode
import googlemaps
from openpyxl import Workbook
import folium
import geopandas as gpd
import ray
from ray import tune
from ray.rllib.agents.ppo import PPOTrainer
from ray.tune.registry import register_env
from gym import Env
from gym.spaces import Discrete, Box
import networkx as nx
import osmnx as ox
```

## 7. Definir a base de dados e a sua análise exploratória.

Para servir de base, usamos todos os dados analíticos de BOs no estado de São Paulo disponível em formato xlsx no site da secretaria de segurança pública. As bases estavam divididas em duas ou mais abas de planilha separando os mesmos dados.



Fizemos uma análise básica da qualidade dos dados antes de começar a EDA. Essa análise consistiu em identificar valores nulos e decidir o que fazer com eles, bem como a padronização de alguns componentes descritivos.

Nos deparamos com dois principais problemas:

1. Falta da informação de latitude e longitude.
  - a. Para este problema, começamos resolvendo ao consultar a API do google maps para inferir a latitude e longitude a partir do logradouro, que é presente em todos os BO's. Essa solução é cara e gastou R\$ 1800,00 em uma rodada que preencheu parcialmente os dados.
  - b. Para os casos que o custo foi um impeditivo, preenchemos os dados com a media de latitude e longitude para o tipo de crime (chamado de rubrica) no bairro em questão. Isso nos permitirá uma simulação mais acurada do que no caso da ausência de valores.
2. Falta da informação acurada de hora e minuto.
  - a. Os dados que vem vazios nesse campo, geralmente vem com uma descrição do horário: "Pela manhã", "A tarde", etc. Preenchemos de acordo com um range a clusterização nos dados que vinham com valores válidos e atribuímos a mediana para a rubrica no bairro em questão. Dessa forma conseguiremos mais insumos na hora da modelagem.

Falando especificamente da EDA, analisamos alguns aspectos para entender o comportamento dos dados:

- **Quantos e quais dados temos a nossa disposição?**
- **Como o crime tem evoluído ao longo do tempo?**
- **Quais os horários mais perigosos na cidade?**
- **Quais os tipos de crimes mais recorrentes?**
- **Quais os locais com mais incidência de crimes?**
- **Quais os bairros mais perigosos?**

## **8. Tratar a base de dados (preparação e treinamento).**

Dividiremos o dataset em dados de tratamento e simulação para que os agentes autônomos pudessem aprender inicialmente a traçar as melhores rotas e também através do retorno das simulações quando conseguiram algum evento de sucesso. Para isso, usamos a divisão aleatória representativa agrupando sempre por duas variáveis relevantes: Bairro e Rúbrica (tipo de crime).

## **9. Definir e descrever as bases teóricas dos métodos analíticos.**

Os métodos analíticos que usaremos serão centralizados na metodologia de MARL em um contexto multiagente, onde:

Os agentes aprendem a tomar decisões para maximizar uma soma cumulativa de recompensas futuras com base nas interações com o ambiente.

Cada agente opera de forma semi-independente, com ações potencialmente afetando outros agentes.

Usaremos os mapas como um grande grafo multidirecional no qual os agentes se movem livremente com uma velocidade constante de 40km/h.

#### **10. Definir e descrever como será calculada a acurácia.**

No contexto de MARL, diretamente, a acurácia é secundária na avaliação da utilidade do modelo. Focaremos de forma intermediária na recompensa cumulativa, convergência e eficiência.

Para avaliar a performance como um todo do projeto, partiremos do pressuposto interpretativo não necessariamente 100% conectado com a realidade de que:

Se o agente estava próximo da ocorrência do crime no momento do crime a situação seria prevenida (inibindo a ação) ou o furto/roubo/ocorrência seria pego em flagrante, levando a uma resolução imediata e isso é considerado um “evento de sucesso”.

A acurácia será calculada no número percentual de “eventos de sucesso” que os agentes conseguiram prever com o aprendizado por reforço.

Como métrica de comparação de eficiência, veremos a comparação desses eventos com os dados de produtividade da polícia, que mostra as apreensões e resoluções dos casos atuais.