

**UNESA Rio de Janeiro**  
**Niterói**

**Análise de Dados sobre a Taxa de Mortalidade no Trânsito no Município do Rio de Janeiro**  
**entre 2000 - 2022**

**Hiago Gonçalves de Souza**  
**Simone Ingrid Monteiro Gama**

**2025.2**  
**Niterói/Rio de Janeiro**

## Sumário

1.	DIAGNÓSTICO E TEORIZAÇÃO.....	3
1.1.	Problemática.....	3
1.2.	Objetivo Principal .....	3
2.	DESENVOLVIMENTO DO PROJETO .....	4
2.1.	Detalhamento técnico do projeto .....	4
3.	ENCERRAMENTO DO PROJETO .....	7
3.1.	Relato de Experiência Individual .....	7
3.1.1.	RESULTADOS E DISCUSSÃO .....	7
3.1.2.	REFLEXÃO APROFUNDADA .....	7
3.1.3.	CONSIDERAÇÕES FINAIS .....	8

## 1. DIAGNÓSTICO E TEORIZAÇÃO

### 1.1. Problemática

O trânsito é uma das principais causas de mortalidade no Brasil, com impacto significativo na saúde pública e na segurança urbana. No estado do Rio de Janeiro, os índices de mortes no trânsito permanecem elevados, mesmo com políticas de fiscalização e campanhas educativas. Essa situação revela a necessidade de compreender, de forma mais detalhada, as tendências e os fatores associados aos óbitos no trânsito, utilizando ferramentas de análise de dados para subsidiar ações de prevenção e políticas públicas mais eficazes.

### 1.2. Objetivo Principal

O objetivo desse trabalho de extensão é apresentar uma análise de dados Big Data sobre a Taxa de Mortalidade no Trânsito no estado do Rio de Janeiro.

## 2. DESENVOLVIMENTO DO PROJETO

### 2.1. Detalhamento técnico do projeto

No início da disciplina extensionista de Tópicos de Big Data 2025.2, foram disponibilizadas diversas bases de dados (Big Data) para escolher aquelas que melhor atendem ao impacto social exigido na disciplina. A equipe deste trabalho escolheu a base de dados sobre a Taxa de Mortalidade no Trânsito no Município do Rio de Janeiro entre 2000 - 2022. ([Link da Base de Dados](#)) ([Link do GitHub com o Projeto](#))

Arquitetura do Sistema:

- Desenvolvido em Python 3.8+
- Utiliza pandas para manipulação de dados
- matplotlib e seaborn para visualizações
- scikit-learn para análises estatísticas

Funcionalidades Implementadas:

1. Leitura inteligente de planilhas Excel com múltiplas abas
2. Limpeza e padronização automática de dados
3. Análise exploratória com histogramas e boxplots
4. Regressão linear para identificação de tendências
5. Geração de relatórios estatísticos completos

### **Distribuição de Frequência – Óbitos (Histograma)**

O histograma apresenta como os valores de óbitos estão distribuídos ao longo dos anos. Ele permite observar faixas em que os valores aparecem com maior frequência.

No gráfico, percebe-se que a maioria dos anos concentra-se entre 700 e 1000 óbitos, indicando que essa faixa representa o comportamento típico do período analisado.

A curva suavizada (KDE) ajuda a visualizar a forma da distribuição e sugere possíveis picos e vales, indicando que há anos com valores significativamente maiores ou menores do que a média geral. Isso revela variações importantes que merecem ser consideradas em análises temporais ou comparativas.

---

### **Evolução Temporal dos Óbitos (Série Temporal)**

Esse gráfico mostra claramente como os óbitos variaram entre 2000 e 2022.

É possível identificar três fases distintas:

- Crescimento inicial (2000–2006): há um aumento consistente nos óbitos, chegando a ultrapassar 1000 mortes anuais.
- Alta variabilidade (2007–2014): o número de mortes oscila bastante, com picos e quedas abruptas, demonstrando instabilidade no cenário da segurança viária.

- Tendência de redução (2015–2022): há um movimento contínuo de queda, chegando aos menores índices da série, especialmente em 2018.

A representação em linha permite visualizar de forma clara o impacto de políticas públicas e eventos externos ao longo dos anos, facilitando a análise de tendências e mudanças estruturais no comportamento dos dados.

## **Boxplot – Óbitos**

O boxplot sintetiza a distribuição dos óbitos por meio de medidas estatísticas. Ele mostra:

- Mediana próxima a 900 óbitos, indicando o valor típico anual.
- Grande amplitude entre os quartis, o que evidencia alta variabilidade ao longo dos anos.
- Valores extremos, acima de 1050 e abaixo de 600, que se destacam como anos atípicos e sugerem impactos pontuais, possivelmente decorrentes de eventos específicos (ações públicas, mudanças nas leis, crises, entre outros).
- O boxplot é útil para identificar a dispersão e detectar assimetrias e outliers, complementando a análise feita pelos outros gráficos.



## **Regressão Linear – Óbitos por Ano:**

O gráfico de regressão linear apresenta a relação entre os anos analisados e o número de óbitos no trânsito, permitindo identificar a tendência geral do comportamento da série histórica. Cada ponto rosa representa os valores observados para determinado ano, enquanto a linha vermelha corresponde ao modelo de regressão linear ajustado aos dados.

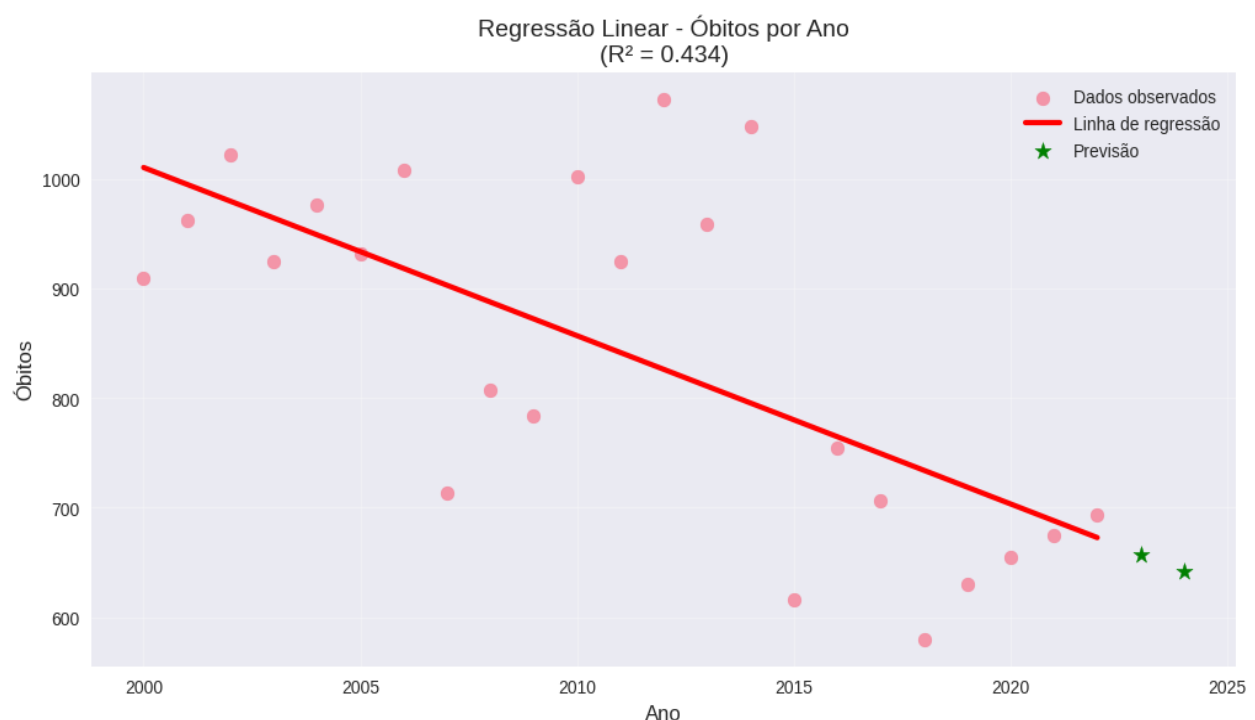
A inclinação negativa da linha evidencia uma **tendência de queda gradual no número de óbitos ao longo do período**. Isso significa que, embora existam oscilações anuais — algumas vezes com picos e outras com reduções acentuadas —, o comportamento geral aponta para uma diminuição consistente das mortes no trânsito no município do Rio de Janeiro entre 2000 e 2022.

O valor apresentado para o coeficiente de determinação ( $R^2 = 0.434$ ) indica que aproximadamente **43,4% da variação dos óbitos pode ser explicada pelo modelo linear**. Embora não seja um valor extremamente alto, ele demonstra que existe uma relação relevante entre o ano e a quantidade de óbitos, mas que outros fatores externos também influenciam significativamente o comportamento da série, tais como:

- mudanças na legislação de trânsito,
- aumento ou diminuição da fiscalização,
- campanhas de conscientização,
- melhorias ou deteriorações na infraestrutura viária,
- crescimento da frota de veículos.

Os pontos verdes ao final da linha representam **previsões geradas pelo modelo de regressão para anos futuros**. Eles sugerem a continuidade da tendência de queda, indicando que, caso as condições históricas se mantenham, o número esperado de óbitos deve permanecer abaixo de 700 nos próximos anos.

Esse tipo de análise é fundamental porque permite não apenas observar o que ocorreu no passado, mas também projetar possíveis cenários futuros, contribuindo para o planejamento e avaliação de políticas públicas voltadas à segurança viária.



### 3. ENCERRAMENTO DO PROJETO

#### 3.1. Relato de Experiência Individual

Nome: Hiago Gonçalves de Souza

Matrícula:202303300743

##### 3.1.1. RESULTADOS E DISCUSSÃO

Ao iniciar o projeto, minha expectativa era compreender melhor como grandes volumes de dados podem ser tratados e utilizados para gerar informações úteis à sociedade. Durante o processo, pude vivenciar na prática a aplicação de conceitos de Big Data, manipulação de bases reais e utilização de bibliotecas específicas para análise estatística.

Observei que os dados sobre mortalidade no trânsito são muito ricos e possibilitam diversas reflexões sobre políticas públicas e comportamentos sociais. Um dos principais resultados obtidos foi a identificação de tendências ao longo dos anos, tanto de crescimento quanto de redução nos óbitos, o que ajuda a entender se as ações governamentais têm surtido efeito.

A experiência me trouxe várias aprendizagens: aprofundamento no uso do Python, maior domínio das bibliotecas de análise de dados, e compreensão mais sólida sobre regressão linear.

Entre as facilidades, destaco o uso das ferramentas de visualização gráfica, que tornam mais clara a interpretação dos dados. Já entre as dificuldades, houve desafios na padronização das planilhas, principalmente devido às diferenças entre as abas e nomes de colunas.

Recomendo que futuros trabalhos desse tipo incluam também análises geográficas ou cruzamento com outras variáveis, como frota de veículos, condições das vias ou investimentos públicos em segurança viária.

##### 3.1.2. REFLEXÃO APROFUNDADA

A experiência vivida dialogou diretamente com a teoria apresentada no relato coletivo e nas aulas. Os conceitos de Big Data, ETL (Extração, Transformação e Carregamento), visualização e modelagem estatística puderam ser experimentados na prática, tornando o aprendizado mais significativo. Percebi que a teoria fornece a base, mas é no desenvolvimento prático, com problemas reais e dados reais, que surgem os desafios e que consolidamos o conhecimento. A possibilidade de aplicar técnicas de regressão linear e ver seus resultados imediatamente reforçou a importância da análise estatística no apoio à tomada de decisões. Além disso, o trabalho reforçou o papel social da extensão universitária, mostrando como a tecnologia pode contribuir para temas relevantes, como mobilidade urbana e segurança no trânsito.

### 3.1.3. CONSIDERAÇÕES FINAIS

O projeto possibilitou compreender de forma profunda como a análise de dados pode apoiar políticas públicas e ações preventivas relacionadas à segurança no trânsito.