

# { ANÁLISE SOBRE PESSOAS DESAPARECIDAS NO MUNICÍPIO DO RJ }



Estácio

TRABALHO EXTENSIONISTA DE TÓPICOS DE BIG DATA 2025.2

Matheus Sant'Ana E Thiago Alves Fraga  
202208858422 202309135159

# OBJETIVO GERAL

REALIZAR UMA ANÁLISE DE DADOS EM BIG DATA SOBRE A POPULAÇÃO DA CIDADE DO RIO DE JANEIRO.

# BASE DE DADOS UTILIZADA

ESCOLHEMOS A BASE DE DADOS :

**PESSOAS DESAPARECIDAS, POR MÊS, TOTAL, DIFERENÇAS COM RELAÇÃO AO ANO ANTERIOR, POPULAÇÃO E TAXA POR CEM MIL HABITANTES, SEGUNDO O ANO, NO MUNICÍPIO DO RIO DE JANEIRO ENTRE JANEIRO/2003 E OUTUBRO/2023**

FONTE: [[HTTPS://WWW.DATA.RIO/DOCUMENTS/EBEF6FA9E5DD4301A0903679AF8ACCCF/ABOUT](https://www.data.rio/documents/EBEF6FA9E5DD4301A0903679AF8ACCCF/ABOUT)]

# TECNOLOGIAS USADAS

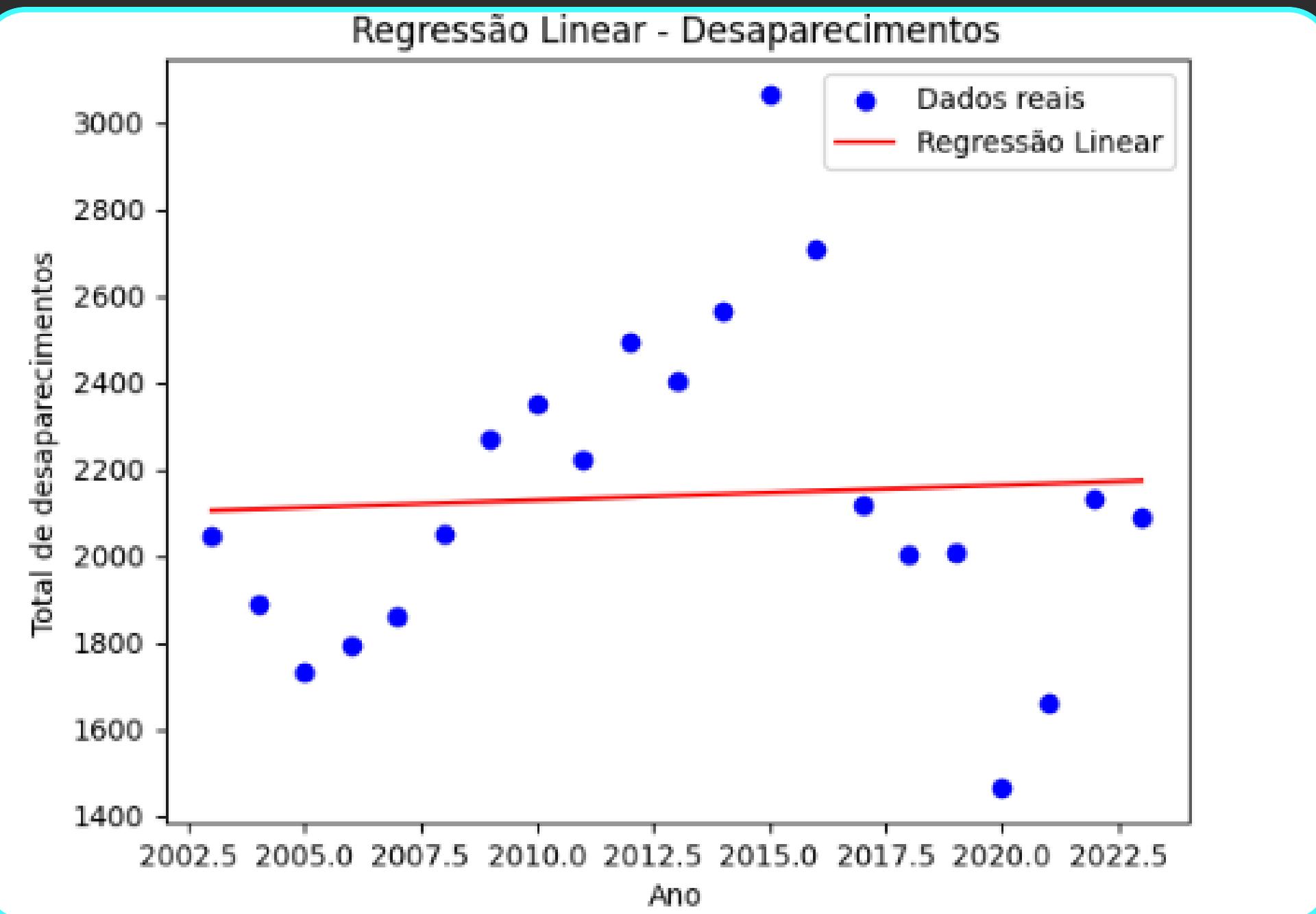
- **Python**: Linguagem de programação central utilizada para todas as etapas de manipulação e análise.
- **Pandas**: Empregada para a leitura, manipulação e limpeza estrutural dos dados, incluindo o tratamento de dados, conversão de tipos de dados e a organização das colunas.
- **PySpark**: Utilizado para o processamento distribuído da base, abordando o volume de dados e o requisito de Big Data de forma eficiente.
- **Scikit-learn** (Sklearn): Aplicado na fase de pré-processamento, realizando a padronização ou normalização de variáveis para garantir a adequação dos dados à modelagem.
- **Matplotlib**: Biblioteca essencial para a visualização gráfica dos dados, como histogramas e gráficos de linha, facilitando a identificação de sazonalidade e tendências de longo prazo.
- **Assistentes de IA**: estruturação e formatação

# DADOS UTILIZADOS

Ano	Pessoas desaparecidas													Total <sup>1</sup>
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez		
2003	167	174	188	155	179	178	137	156	210	165	180	155		2 044
2004	185	165	177	175	170	128	133	138	183	149	133	153		1 889
2005	177	119	157	127	147	142	126	157	118	185	124	153		1 732
2006	178	148	150	165	136	136	139	148	145	165	122	160		1 792
2007	162	139	176	160	140	187	170	162	146	150	133	133		1 858
2008	157	150	147	147	131	150	148	226	218	199	206	171		2 050
2009	194	187	223	181	157	177	202	173	182	184	194	214		2 268
2010	218	212	190	180	177	192	160	180	201	223	193	224		2 350
2011	236	198	208	183	176	202	181	171	154	171	168	173		2 221
2012	213	214	187	180	214	198	229	209	188	247	185	232		2 496
2013	226	203	238	237	205	214	137	177	182	214	168	203		2 404
2014	254	209	230	227	164	182	166	192	184	222	280	256		2 566
2015	266	232	286	249	255	238	237	252	267	251	276	257		3 066
2016	265	249	235	223	230	179	241	201	215	254	216	202		2 710
2017	150	118	162	158	183	177	176	196	210	239	183	165		2 117
2018	199	144	203	157	162	171	136	175	142	183	170	163		2 005
2019	201	151	168	153	173	157	165	155	171	190	165	160		2 009
2020	187	158	110	81	94	85	98	130	141	117	130	134		1 465
2021	164	125	133	109	111	127	123	128	150	168	149	173		1 660
2022	176	156	170	167	159	175	182	224	164	201	165	192		2 131
2023	228	223	258	204	194	171	206	196	224	183	..	..		2 087

# E QUAIS FORAM OS RESULTADOS?

# REGRESSÃO LINEAR



- Inclinação: 3.45...
- Intercepto: -4814.95...

- Em resumo, a regressão linear sugere que, embora haja flutuações anuais, a tendência geral é de um ligeiro aumento no número total de desaparecimentos por ano, de acordo com os dados fornecidos.

	Frequência Absoluta	Frequência Relativa (%)
Jan	4203.0	9.36
Fev	3674.0	8.18
Mar	3996.0	8.90
Abr	3618.0	8.05
Mai	3557.0	7.92
Jun	3566.0	7.94
Jul	3492.0	7.77
Ago	3746.0	8.34
Set	3795.0	8.45
Out	4060.0	9.04
Nov	3540.0	7.88
Dez	3673.0	8.18

## FREQ. ABSOLUTA

- A frequência absoluta mostra o número total de desaparecimentos registrados para cada mês ao longo de todos os anos presentes no conjunto de dados.
- A frequência relativa nos ajuda a ver a importância de cada mês em termos de desaparecimentos em comparação com os outros meses, mostrando quais meses têm uma proporção maior ou menor de ocorrências.
- O mês de Janeiro segue sendo o mês com maiores casos de desaparecimentos por ano, seguido do mês de Outubro e Março.

# CONCLUSÃO

Ao analisarmos a regressão linear do "Ano" versus "Total de desaparecimentos", podemos aprender sobre a tendência temporal, mas também percebemos que outros fatores, como a população, podem ser mais influentes e merecem ser investigados em análises futuras para obter uma compreensão mais completa.

**OBRIGADO!**