

Senac

Todas as formas  
de aprender

tecno  
logia

**CURSO**

**\_BIG  
DATA  
SCIENCE**

**\_ANALISTA DE DADOS**



**ALESSANDRO VIEIRA**



**\_SEJA MUITO  
BEM-VINDO  
À NOSSA TRILHA!**



Unidade Curricular	Carga Horária
<b>UC1:</b> Desenvolver algoritmos, versionamentos e linguagem de consulta estruturada.	44 horas
<b>UC2:</b> Criar e manipular dados utilizando matemática estatística.	76 horas
<b>Carga Horária Total</b>	<b>120 horas</b>



# APRESENTAÇÃO E **CONTEXTUALIZAÇÃO**



## Objetivo do curso

Subsidiar o participante no exercício do pensamento analítico e na interpretação de um conjunto de dados, possibilitando a manipulação de grandes volumes, gerando e apresentando análises e insights que tragam valor para o negócio.

## Indicadores da UC2

Desenvolve análise estatística de dados de acordo com os requisitos propostos.

Manipula base de dados, utilizando a linguagem Python e suas bibliotecas.

Constrói dashboard com Power Bi, fornecendo uma representação visual.

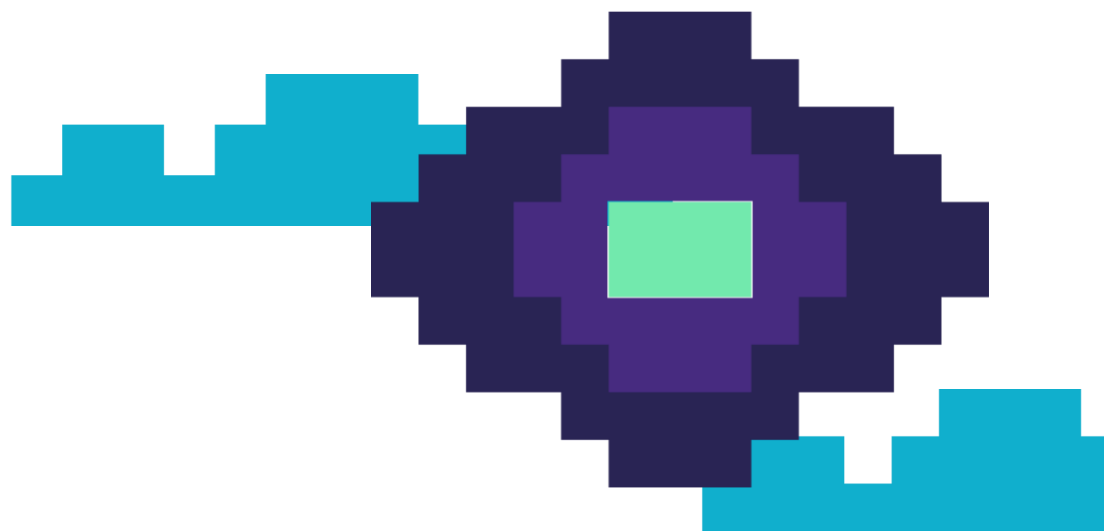


# BORA COMEÇAR?

AULA

01

# Introdução à Estatística







## CENÁRIO



# ANALISTA DE DADOS – BIG DATA SCIENCE

## Índice de Criminalidade no Estado do Rio de Janeiro

Foi aberta uma licitação no Estado do Rio de Janeiro para a contratação de uma consultoria que desenvolva projetos de business intelligence para gestão dos indicadores de criminalidade na cidade. Esse trabalho será apresentado para o Secretário de Segurança Pública e o objetivo é dar subsídios para a tomada de decisão data-oriented e criar estratégias de segurança relacionadas a todo o tipo de criminalidade.

O secretário de segurança pública tem como objetivo apresentar à mídia um indicador positivo na redução de crimes, com base no histórico dos últimos anos.



# CENÁRIO



## ANALISTA DE DADOS – BIG DATA SCIENCE

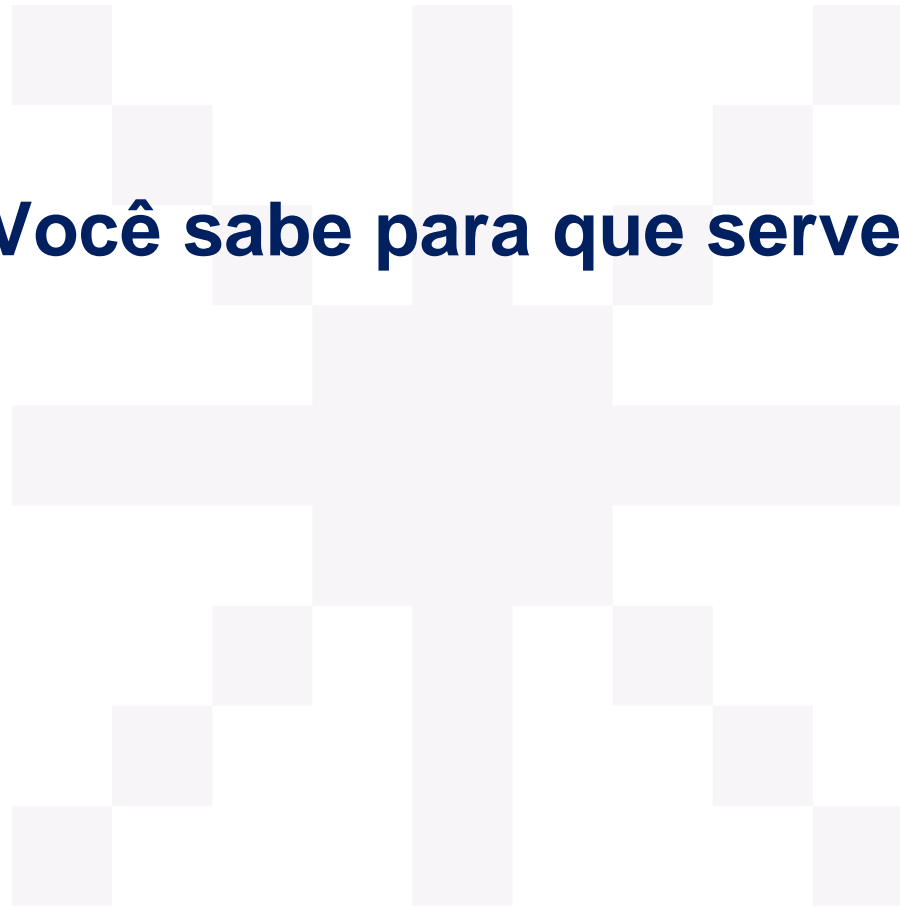
### Índice de Criminalidade no Estado do Rio de Janeiro

Você deverá realizar um estudo dos dados baseado em análise descritiva, a fim dar subsídios para que a equipe proponha uma nova Política de Segurança Pública para diminuir os índices de criminalidade.

**E você? Está pronto para esse desafio?**



# Você sabe para que serve a Estatística?



# ANALISTA DE DADOS – BIG DATA SCIENCE

A **Estatística** é a ciência que envolve a coleta, análise, interpretação e apresentação de dados.

Utilizada em diversas áreas, como saúde, economia e ciências sociais.

As Medidas de **tendência central** e **dispersão** são ferramentas essenciais para resumir e interpretar conjuntos de dados.

# ANALISTA DE DADOS – BIG DATA SCIENCE

As medidas de **tendência central** são usadas para identificar um valor que representa o centro de um conjunto de dados.

Principais medidas:

- **Média Aritmética:** soma dos valores dividida pela quantidade de dados.
- **Mediana:** valor que ocupa a posição central quando os dados estão ordenados.
- **Moda:** valor que mais aparece em um conjunto de dados.

# ANALISTA DE DADOS – BIG DATA SCIENCE

**Média Aritmética** : soma dos valores dividida pela quantidade de dados.

Conjunto de dados: 5, 8, 10, 6, 7

$$\text{Média} = (5 + 8 + 10 + 6 + 7) / 5 = 36 / 5 = \mathbf{7,2}$$

# ANALISTA DE DADOS – BIG DATA SCIENCE

- **Mediana:** valor que ocupa a posição central quando os dados estão ordenados.

Conjunto de dados: 3, 5, 7, 9, 11

Mediana = **7** (valor central)

***OBS: Se o número de dados for par, a mediana será a média dos dois valores centrais.***

Conjunto de dados: 3, 5, 7, 8, 9, 11

Mediana =  $(7 + 8) / 2 = 15 / 2 = \mathbf{7,5}$

# ANALISTA DE DADOS – BIG DATA SCIENCE

- **Moda:** valor(es) que mais aparece(m) em um conjunto de dados.

Conjunto de dados: 4, 5, 6, 6, 7

Moda = **6**

Conjunto de dados: 4, 5, 6, 6, 7, 8, 9, 9

Moda = **6 e 9 (Bimodal)**



# ANALISTA DE DADOS – BIG DATA SCIENCE

As medidas de **dispersão** indicam o grau de variação dos dados em relação à média.

Principais medidas:

- **Amplitude**: diferença entre o maior e o menor valor.
- **Variância**: média dos quadrados dos desvios em relação à média.
- **Desvio Padrão**: raiz quadrada da variância.

# ANALISTA DE DADOS – BIG DATA SCIENCE

- **Amplitude:** diferença entre o maior e o menor valor.

Conjunto de dados: 4, 7, 10, 15

$$\text{Amplitude} = 15 - 4 = 11$$

# ANALISTA DE DADOS – BIG DATA SCIENCE

- **Variância:** média dos quadrados dos desvios em relação à média.

Conjunto de dados: 4, 8, 6

$$\text{Média} = (4 + 6 + 8) / 3 = 18 / 3 = \mathbf{6}$$

$$\text{Var} = [(4 - 6)^2 + (8 - 6)^2 + (6 - 6)^2] / 3 = (4 + 4 + 0) / 3 = \mathbf{2,67}$$

# ANALISTA DE DADOS – BIG DATA SCIENCE

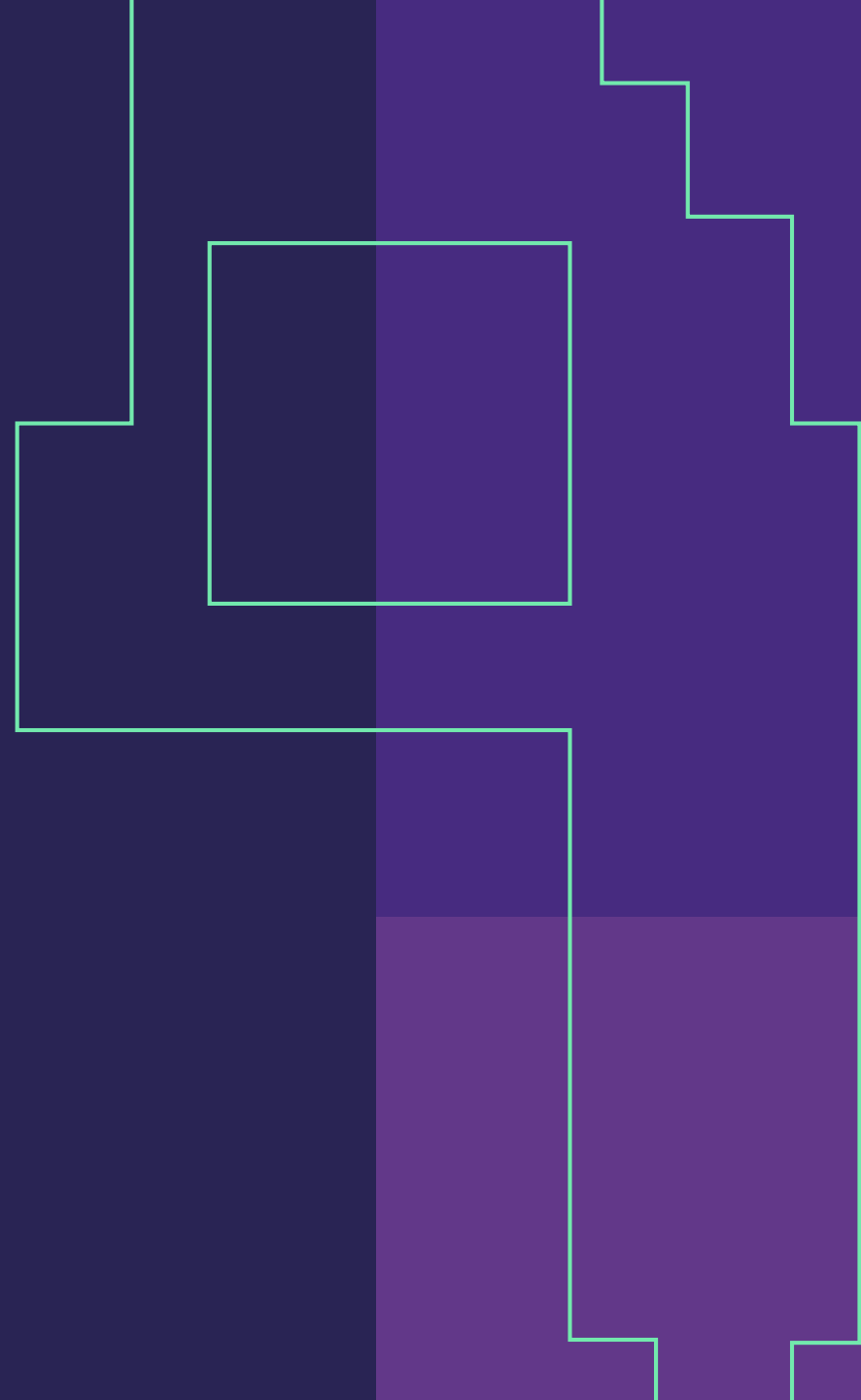
- **Desvio Padrão:** raiz quadrada da variância.

Variância = 2,67

Desvio Padrão =  $\sqrt{2,67} \approx 1,63$

A cluster of five small purple squares arranged in a cross-like pattern around the text.

# BORA **PRATICAR** ?



# ANALISTA DE DADOS – BIG DATA SCIENCE

## Praticando

1- Calcule a média, mediana e moda do conjunto de dados: 10, 20, 20, 30, 40.

2- Calcule a variância e o desvio padrão do conjunto de dados: 5, 10, 15.

3- Um estudante fez algumas provas em seu curso e obteve as notas 13, 34, 45, 26, 19, 27, 50, 63, 81, 76, 52, 86, 92 e 98. Determine a sua média. Essa medida é ideal para representar o seu desempenho? Justifique sua resposta.

# ANALISTA DE DADOS – BIG DATA SCIENCE

## Praticando

4- Considere a distribuição a seguir relativa a notas de dois alunos de matemática durante determinado semestre:

Aluno A	9,5	9,0	2,0	6,0	6,5	3,0	7,0	2,0
Aluno B	5,0	5,5	4,5	6,0	5,5	5,0	4,5	4,0

a) Calcule as notas médias de cada aluno.

b) Qual aluno apresentou resultado mais homogêneo? Justifique.

**Chegamos ao  
final do nosso  
encontro.**

Até o próximo.



Não esqueça de reservar um momento do seu dia  
para fazer a **JORNADA DE FIXAÇÃO**.