

---

**GEIA 25.1- INTRODUÇÃO A CIÊNCIA DE DADOS**

---

**Aluno:**

SEU NOME COMPLETO (MATRÍCULA INSTITUCIONAL), ex: Fulano da Silva  
(1234567)

---

**Monitor:**

Arthur Veras

---

## Desafio – Análise Estatística de Dados de Alunos

### Descrição do Desafio:

Você será responsável por criar um programa em Python que realize uma **análise estatística completa** sobre um conjunto de dados de alunos de um grupo de estudos fictício. O seu programa deve utilizar os conceitos de **estatística descritiva** e **identificação de outliers** estudados em aula.

Seu sistema deverá ser capaz de:

- **Importar Dados:** Receber como entrada um arquivo `.csv` contendo dados de vendas. O arquivo terá colunas como: `ID`, `Idade`, `Nota_Matematica`, `Nota_Estatistica`, `Nota_Programacao`, `Faltas`, `Sexo` e `Cidade`.
  - **Análise Estatística:**
    - Calcular e exibir: **Média**, **Moda**, **Mediana**, **Desvio Padrão** e **Variância**, para todas as colunas numéricas.
  - **Detecção de Outliers:**
    - Implementar a identificação de outliers utilizando **Z-Score** e **Cerca de Tukey** para os campos de notas e faltas.
    - Gerar uma listagem separada dos registros considerados outliers.
  - **Relatório Final:**
    - Gerar um resumo no terminal, contendo:
      - Principais medidas estatísticas.
      - Quantidade de outliers detectados por cada método (Z-Score e Cerca de Tukey).
      - Lista dos IDs das notas de alunos identificados como outliers, em **matemática**, **programação** e **estatística**.
-

## Requisitos Técnicos:

### 1. Organização de Código:

- Criar pelo menos três classes:
  - `ImportarDados`: Para importar e organizar os dados.
  - `AnaliseEstatistica`: Para realizar os cálculos estatísticos.
  - `DeteccaoOutliers`: Para aplicar Z-Score e Cerca de Tukey.

### 2. Métodos Obrigatórios:

- `media()`, `mode()`, `median()`, `std()`, `interval()`, `variance()`, `zscore()`, `cerca_tukey()`, `generate_report()`.

3. **Dado a ser Utilizado:** A base de dados que vocês vão utilizar neste processo vai estar no nosso repositório dentro da pasta [data/](#)

**PS: Quaisquer interação(input ou output) deve ser feita via terminal.**

---

## Critérios de Avaliação:

O trabalho será avaliado com base nos seguintes critérios:

### 1. Cumprimento dos Requisitos (30 pontos):

Implementação correta das funcionalidades solicitadas.

### 2. Organização e Clareza do Código (25 pontos):

Código limpo, organizado, bem comentado e de fácil compreensão.

### 3. Aplicação Correta de Conceitos de Estatística e POO (25 pontos):

Boa prática de conceitos como encapsulamento, separação de responsabilidades e correta aplicação dos cálculos estatísticos.

### 4. Documentação (20 pontos):

Explicação clara no próprio código (comentários) e documentação do código usando esse template também será levada em consideração, o template foi disponibilizado no nosso grupo do **WhatsApp**.

---

## Importante:

- **Originalidade:** O uso de **códigos prontos** (de fontes externas), **compartilhamentos de códigos entre os alunos** ou a **aplicação de inteligência artificial**, como resposta automática, **resultará em nota zero**. O trabalho deve ser feito de forma independente, com o código escrito pelo aluno.
  - Bibliotecas permitidas: **Math, Numpy, Csv**.
  - **Entrega:** A entrega deve ser feita via **GitHub** do grupo de estudo, seguindo as regras pré-estabelecidas.
- 

## Conclusão:

Este desafio tem como objetivo reforçar os conhecimentos de **estatística descritiva** e **visualização de dados**, aplicando os conceitos estudados de maneira prática. Ao desenvolver o trabalho, o aluno poderá exercitar a interpretação de dados, o **cálculo correto de medidas estatísticas** e a **identificação de outliers**, além de representar essas informações de forma clara e organizada. A atividade também busca incentivar a autonomia, a originalidade e a boa organização do raciocínio, habilidades essenciais para quem deseja seguir na área de Ciência de Dados. **Boa sorte a todos!**