



Universidade Federal do Ceará  
Centro de Ciências  
Departamento de Computação  
Bacharelado em Engenharia de Computação

## **CK00509 – Mineração e Ciência de Dados**

### **2022.1**

### **Trabalho 2: Análise Exploratória dos Dados**

#### **1. Descrição:**

Este trabalho consiste na utilização de métodos estatísticos com a finalidade de explorar e extrair conhecimento do dataset gerado no trabalho anterior.

#### **2. Técnicas a serem utilizadas:**

##### **1. Métodos de Análise Estatística dos Dados:**

- a) Medidas de Tendência Central
- b) Medidas de Variabilidade
- c) Tabela de Frequência e Histograma
- d) Boxplot
- e) QQ-Plot
- f) Correlações
- g) Teste de Normalidade
- h) Best Fit Distribution

#### **3. Avaliação**

Espera-se com a realização deste trabalho que cada estudante elabore e entregue (de forma digital) os seguintes documentos:

- Arquivo CSV contendo os dados do Dataset;
- Jupyter Notebook contendo as análises e visualizações realizadas sobre os dados.
- Vídeo (disponibilizado no Youtube) apresentando e descrevendo as atividades desenvolvidas.

A avaliação deste trabalho se dará em duas etapas:

1ª. Vídeo de Apresentação do Estudo Exploratório: Cada estudante irá disponibilizar um vídeo (no Youtube) apresentando o dataset gerado e o processo utilizado para explorar os dados. O estudante pode utilizar slides e notebooks.

2ª. Avaliação da Análise: O professor da disciplina irá avaliar a qualidade das análises realizadas pelo estudante, bem como dos códigos implementados.

A avaliação do trabalho irá envolver os seguintes quesitos:

- Abrangência e Organização do Notebook
- Qualidade dos Códigos Utilizados
- Clareza do Texto Utilizado para Descrever as Análises
- Domínio do Tema
- Relacionamento do Tema com a Área de Segurança (desejável, mas não obrigatório)

#### **4. Data da Entrega: 15/07/2022**

- PS. Não serão aceitos trabalhos que não forem apresentados (por meio de vídeo disponibilizado no Youtube).
- PS. Cada estudante será responsável pela disponibilização do ambiente (software e hardware) necessário para a gravação da apresentação do seu trabalho.
- Os Notebooks e Arquivos CSV deverão ser disponibilizados, em formato .ZIP, em uma das seguintes plataformas: SIGAA, Classroom ou em um repositório público (GitHub ou GitLab).
- Utilize o formulário a seguir para indicar as URLs do vídeo e do arquivo ZIP contendo os códigos e os dados utilizados:

<https://forms.gle/k23yymu2dQUPG8gv7>

“Criar o que não existe ainda deve ser a pretensão de todo sujeito que está vivo”.

**Paulo Freire**