

# Ciência de Dados - 2024/2: Trabalho I

## Análise de um *dataset*

Cada grupo deve trabalhar no dataset escolhido e apresentar as etapas no *Google Colab*, documentando com clareza os passos e explicações necessárias. Roteiro:

### 1) Caracterização do Dataset

- **Descrição Geral:** Apresentar um resumo do dataset, incluindo:
  - Origem e contexto do dataset (Referencia)
  - Características principais (colunas) e seus significados.
  - Classe alvo (se aplicável), variável preditora principal.
- **Estatísticas Básicas:** Análise inicial de estatísticas descritivas:
  - Quantidade de linhas (observações) e colunas (atributos).
  - Média, mediana, moda, mínimo, máximo.
  - Desvios padrão, correlação entre variáveis e distribuições.
  - Verificação de valores ausentes e outliers.

### 2) Objetivo de Análise

- **Objetivo:** Definir o problema ou questão que se deseja investigar com o dataset.
- **Ferramentas e Técnicas de Análise:** Escolher a técnica mais adequada para atingir o objetivo entre:
  - **Regressão Linear:** Caso o objetivo seja prever valores contínuos.
  - **k-Nearest Neighbors (k-NN):** Para tarefas de classificação ou predição baseada em proximidade.
  - **Regressão Logística:** Se o objetivo envolve prever categorias binárias ou multiclases.
  - **Análise de Componentes Principais (PCA):** Para reduzir a dimensionalidade e visualizar a estrutura do dataset.

### 3) Resultados

- **Execução da Técnica Escolhida:**
  - Implementar a técnica no Colab.
  - Explicar as etapas, justificando parâmetros escolhidos.
- **Visualização dos Resultados:**
  - Incluir gráficos e figuras relevantes.
  - Se possível, interpretar cada figura; o que revelam sobre os dados.

### 4) Conclusões e Aplicabilidade

- **Conclusão da Análise:** Resumo dos achados principais e se o objetivo foi atingido.
- **Aplicabilidade das Técnicas:** Discutir a adequação da técnica escolhida para o problema, seus pontos fortes e limitações.
- **Sugestões para Análises Futuras:** o que poderia ser melhorado ou aprofundado. O plano de trabalho poderia ser utilizado em outros conjuntos?

Apos finalizado o trabalho, verificar se o notebook está bem organizado, com uso de markdown para títulos, descrições e explicações. Cada etapa deve estar bem documentada para que qualquer pessoa possa entender o processo.

### Apresentação em 5'

A primeira parte da aula do dia ~~5/11~~<sup>12/11</sup> será destinada a apresentação dos trabalhos pelos grupos. O notebook será mostrado com o *datashow*: um integrante controla e roda as fases em tanto o outro explica.