



CURSO | TECNOLOGIA EM CIÊNCIA DE DADOS

Professor

Carolina Toledo Ferraz

Componente

Análise exploratória de dados

Aula

Aula 7

Aplicando conhecimento

Projeto de Análise Exploratória de Dados (EDA)

Aplicar as técnicas de Análise Exploratória de Dados (EDA) estudadas durante o curso para investigar e extrair *insights* de um conjunto de dados real, utilizando o R ou o Python.

O aluno deverá utilizar seleções, resumos estatísticos, gráficos e transformações para responder a perguntas sobre os dados, levantar hipóteses e apresentar uma conclusão final.

Descrição do Projeto:

1. Escolha de Base de Dados:

- Você pode usar uma base pública, desde que tenha:
 - Pelo menos 4 variáveis numéricas e/ou categóricas;
 - Pelo menos 50 registros.

2. Etapas obrigatórias que deverão realizar:

2.1. Exploração Inicial dos Dados:

- Descrever as variáveis presentes.
- Verificar se há dados ausentes ou valores discrepantes (*outliers*).

2.2. Resumo Estatístico:

- Calcular medidas como média, mediana, moda, mínimo, máximo, quartis e desvio padrão para variáveis numéricas.

2.3. Visualização de Dados:



- Criar pelo menos 3 gráficos, escolhendo entre:
 - Histogramas
 - Boxplots
 - Gráficos de dispersão (scatter plot)
 - Mapas de calor (Heatmaps)

2.4. Teste de Normalidade (opcional para desafio extra):

- Realizar teste Kolmogorov-Smirnov para alguma variável numérica.

2.5. Relações Entre Variáveis:

- Investigar se há correlação entre duas variáveis numéricas.

2.6. Conclusão:

- Relatar os principais padrões, anomalias e relações identificadas.
- Discutir *insights* relevantes baseados nas análises.

Entrega:

- Um arquivo .ipynb (Python - Jupyter) ou .R
 - Código completo,
 - Gráficos,
 - Análises interpretadas.
- E um pequeno **relatório final** (máximo 2 páginas em pdf) com resumo das descobertas.