# **Projeto da UC Análises Descritivas e Modelos Analíticos para problemas de negócios**

O trabalho deverá ser realizado em grupos de até 5 pessoas. A entrega do trabalho deverá ocorrer até o dia 10/06 às 23:59 e deverá ser um link do GitHub com o projeto.

O projeto é uma análise de dados que deve conter as seguintes etapas:

* Carregar e ler arquivos .csv, xlsx ou de um banco de dados;
* Avaliar a necessidade da transformação da variável em outra escala (agrupar idade por faixas, por exemplo);
* Realizar a codificação das variáveis categóricas de acordo com os valores das variáveis (label encoder, one hot encoder ou target encoder);
* Normalizar as variáveis;
* Avaliar a necessidade de realizar o balanceamento da variável alvo;
* Tratar variáveis com alta correlação;
* Realizar a seleção de variáveis;
* Aplicar um modelo de regressão ou classificação utilizando uma técnica de hiperparametrização automática;
* Aplicar uma medida de avaliação do modelo.
* Gerar uma visualização para o resultado de acordo com os tipos abaixo:
  + Agrupamento: gerar a visualização Scatter;
  + Classificação: gerar a matriz de confusão;
  + Regressão: gerar a visualização da linha de saída.

Descritivo das pontuações:

* 1 ponto pela implementação e explicação do uso de cada acima (ou do não uso quando não for viável);

Enviar o projeto e o link do github no email [paulo.nietto@animaeducacao.com.br](mailto:paulo.nietto@animaeducacao.com.br).

Informar os membros dos grupos em: https://forms.office.com/r/ti6DB89vTQ