

CURSO SISTEMAS DE INFORMAÇÃO

MINERAÇÃO DE DADOS

ATIVIDADE

Cada aluno deverá aplicar técnicas de pré-processamento e transformação em um dataset real, diferente dos colegas, explorando problemas variados de qualidade de dados e preparando-os para análises mais avançadas.

Datasets Disponíveis (Kaggle)

- Movies Dataset – <https://www.kaggle.com/datasets/bharatnatrayn/movies-dataset-for-feature-extracion-prediction?select=movies.csv>
- Food Choices – https://www.kaggle.com/datasets/borapajo/food-choices?select=food_coded.csv
- Data Science Job Postings – https://www.kaggle.com/datasets/rashikrahmanpritom/data-science-job-posting-on-glassdoor?select=Uncleaned_DS_jobs.csv
- Audible Dataset – https://www.kaggle.com/datasets/snehangsude/audible-dataset?select=audible_uncleaned.csv
- Mexican Federal Government Salaries – <https://www.kaggle.com/datasets/ivansabik/mexican-federal-government-salaries>

Distribuição dos Datasets por Aluno

- ALFREDO DE SOUZA AGUIAR NETO → Movies Dataset
- ÁLVARO CALEBE SOUZA FERREIRA → Food Choices
- ANA BEATRIZ MARQUES MOREIRA → Data Science Job Postings
- CAMILLE GOMES DE MENEZES → Audible Dataset
- CHARLES ABRAÃO PANTOJA SANTOS → Mexican Salaries
- FELIPE FERREIRA SOUSA → Movies Dataset

- FILIPE DIAS BARRETO → Food Choices
- GABRIEL ALVES DA SILVA → Data Science Job Postings
- GUILHERME SILVEIRA FIGUEIRA → Audible Dataset
- HEMILLY CHRISTINNE SILVA PEREIRA → Mexican Salaries
- HENRIQUE PONTES BARROS AMORIM → Movies Dataset
- ISAC PEREIRA DA SILVA → Food Choices
- JOANA LETÍCIA AZEVEDO DO NASCIMENTO SOARES → Data Science Job Postings
- JOÃO VICTOR DE OLIVEIRA CARNEIRO → Audible Dataset
- JOÃO VICTOR FERNANDES CORDEIRO MARTINS → Mexican Salaries
- JOÃO VICTOR PÓVOA FRANÇA → Movies Dataset
- JÚLIA PEREIRA GONÇALVES → Food Choices
- KAIO MACEDO MACHADO → Data Science Job Postings
- KATRINA DE MORAIS SANTOS → Audible Dataset
- LEANDRO GUIMARÃES GARCIA → Mexican Salaries
- LUCAS ARAUJO DE CASTRO → Movies Dataset
- LUCAS DANIEL RODRIGUES DOS SANTOS → Food Choices
- LUCAS HENDERSON VIEIRA CUNHA → Data Science Job Postings
- LUÍS EDUARDO NOLETO DE MELO → Audible Dataset
- MARCELO ALVES LEITE → Mexican Salaries
- MATEUS TOSHIO COLI SHIBATA → Movies Dataset
- OTAVIO NARDINI DE FARIA → Food Choices
- VICTOR ALVES DOS REIS DIAS → Data Science Job Postings

Tarefas Obrigatórias

1. Carregar e explorar o dataset (head, info, describe).
2. Limpeza de dados:
 - Tratar valores ausentes.
 - Remover duplicatas.
 - Padronizar formatos (datas, textos, unidades).
3. Transformações:
 - Normalização ou padronização (justifique a escolha).
 - Transformação logarítmica (se necessário).
 - Extração de atributos de data (ano, mês, dia).
 - Encoding de variáveis categóricas (One-Hot ou Label, conforme adequado).
4. Visualização exploratória:
 - Histograma, boxplot, scatterplot e gráficos de contagem.
5. Resumo escrito: explique como cada transformação melhorou a qualidade ou utilidade dos dados.

Entrega

Notebook Jupyter (.ipynb) com código bem comentado, gráficos e análise.