

PROJETO FINAL

Trabalho em grupo: 3 ou 4 elementos (preferencialmente de 4)

Ferramentas: Orange Data Mining + Python

Objetivo: Aplicar técnicas de *clustering*, deteção de anomalias e recomendação, integrando visualização e interpretação clara dos resultados.

Cada grupo pode escolher uma das propostas.

PROPOSTA 1 - ANÁLISE INTELIGENTE DE DADOS URBANOS

1. Descrição

O objetivo deste projeto é desenvolver uma aplicação simples que analise dados relacionados com uma cidade, como ruído, trânsito, poluição, clima ou sensores urbanos, e desenvolva um sistema com:

- Clusters representando padrões urbanos
- Deteção de anomalias (picos ou eventos estranhos)
- Recomendações simples e úteis
- Visualização clara dos resultados obtidos

O enfoque é prático e interpretativo, não matemático.

2. Tarefas Obrigatórias

2.1 Clustering, no Orange

- Carregar *dataset* escolhido
- Aplicar **K-Means** ou outro método disponível
- Identificar 2 a 4 clusters
- Interpretar cada cluster (exemplo.: “horários com ruído elevado”)
- Incluir captura do fluxo do Orange

2.2 Deteção de Anomalias - Python ou Orange

Técnicas permitidas (escolher pelo menos uma): Método Gaussiano, Z-score, Isolation Forest (opcional e simples), Widget de outliers no Orange

A interpretação é obrigatória (exemplo.: “pico de ruído às 23h → evento incomum”).

2.3 Sistema de Recomendação

Criar recomendações baseadas nos clusters e anomalias.

Exemplos:

```
if ruido > limite:  
    recomendar("Evitar esta zona agora.")  
  
if cluster == "trânsito alto":  
    recomendar("Usar transporte público neste período.")
```

O objetivo é **gerar recomendações úteis**, mesmo que simples.

2.4 Aplicação / Dashboard

O grupo pode escolher:

- Jupyter Notebook bem organizado
- Interface Tkinter simples (botões, gráficos e recomendações)
- Script Python + relatório descritivo

3. Bases de Dados Recomendadas (com links)

(A) *CityPulse Dataset* – dados urbanos (ruído, mobilidade, sensores)

<http://iot.ee.surrey.ac.uk:8080/datasets.html>

(B) *UCI Air Quality Dataset* – qualidade do ar

<https://archive.ics.uci.edu/dataset/360/air+quality>

(C) *Bike Sharing Dataset* (UCI / mobilidade urbana)

<https://archive.ics.uci.edu/dataset/275/bike+sharing+dataset>

PROPOSTA 2 - ANÁLISE INTELIGENTE DO CONSUMO ELÉTRICO

1. Descrição

O objetivo desta proposta é criar uma aplicação que analise dados de consumo energético numa casa e produza:

- Clusters de perfis de consumo
- Detecção de picos e comportamentos anómalos
- Recomendações práticas para poupança
- Visualização simples e interpretação clara

2. Tarefas Obrigatórias

2.1 Clustering - Orange

- Usar K-Means ou semelhante
- Identificar perfis como:
 - “Casa com consumo noturno elevado”
 - “Casa com consumo regular”
- Exportar gráfico e fluxo Orange

2.2 Detecção de Anomalias

Técnicas possíveis: Z-Score, Gaussian anomaly detection, Outliers no Orange, simples análise de limites (exemplo *threshold*)

Exemplos:

- pico às 3h → possível eletrodoméstico a funcionar indevidamente
- valores muito acima da média → alerta

2.3 Sistema de Recomendação

Regras simples, como:

```
if consumo_nocturno > media:
    recomendar("Usar máquina de lavar durante o dia.")
```

```
if anomalia_detectada:
    recomendar("Verificar aquecedor ou frigorífico.")
```

O foco é a lógica, não a complexidade.

2.4 Aplicação / Dashboard

O formato é livre, podendo ser:

- Jupyter Notebook bem organizado
- Interface Tkinter simples (botões, gráficos e recomendações)
- Script Python + relatório descritivo

3. Bases de Dados Recomendadas (com links)

(A) *Household Electric Power Consumption* (UCI)

<https://archive.ics.uci.edu/dataset/235/individual+household+electric+power+consumption>

(B) *Smart Home Dataset* (Kaggle)

<https://www.kaggle.com/datasets/taranvee/smart-home-dataset-with-weather-information>

(C) *Electricity Load Diagrams Data Set* (UCI)

<https://archive.ics.uci.edu/dataset/321/electricity+load+diagrams+20112014>

ENTREGAS OBRIGATÓRIOS

1. Relatório final (6–10 páginas)

Deve incluir:

- Introdução ao problema
- Descrição do dataset
- Fluxo do Orange (com imagens)
- Resultados dos clusters
- Anomalias detetadas
- Regras de recomendação
- Conclusões e limitações

2. Pasta com código fonte

- Python organizado
- ficheiros .ows do Orange
- dataset usado

3. Defesa – Obrigatória – Sem defesa o projeto é desconsiderado

- Demonstração da aplicação
- Explicação do processo

Avaliação

Critério	Valores
Clustering (Orange)	2
Deteção de Anomalias	2
Sistema de Recomendação	2
Organização do código	2
Relatório	4
Defesa	8