

# Trabalho prático ICD 2023 2024

Neste repositório encontra-se o trabalho prático da unidade curricular Introdução às Ciências de Dados.

## first-sprint

Nesta pasta foi realizado o pre-processamento dos dados, a análise exploratória e a criação do primeiro modelo LDA.

O ficheiro central do projeto é o main.ipynb, que abriga o código fundamental do trabalho e é estruturado como um Jupyter Notebook. Para garantir uma execução eficiente e isolada, optamos pela criação de um ambiente virtual, no qual todas as dependências necessárias, detalhadamente listadas no ficheiro requirements.txt, foram meticulosamente instaladas. Essa estratégia assegura não apenas a organização e a portabilidade do projeto, mas também a sua compatibilidade e funcionamento otimizado.

Também existem outros ficheiros neste repositório como os ficheiros com os diferentes dados sobre os artigos.

## graps

Nesta pasta encontra-se a interface que foi realizada para o projeto no ficheiro our\_app.py.

Este ficheiro pode ser corrido pelo terminal com o comando:

```
cd graphs  
streamlit run our_app.py
```

Neste folder também estão alguns ficheiros com os dados que foram utilizados para a realização dos gráficos. E o ficheiro main2.py que tem uma implementação do modelo LDA que tem como input um país e é utilizado diretamente pelo ficheiro our\_app.py.