



ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO  
Licenciatura em Engenharia Informática

SISTEMAS DE INFORMAÇÃO

# EXPLORAÇÃO E DESIGUALDADE LABORAL GLOBAL ATRAVÉS DE DADOS ABERTOS

Engenharia de Dados

João Augusto Costa Branco Marado Torres



Beja, novembro de 2025

INSTITUTO POLITÉCNICO DE BEJA  
ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO  
Licenciatura em Engenharia Informática  
SISTEMAS DE INFORMAÇÃO

# EXPLORAÇÃO E DESIGUALDADE LABORAL GLOBAL ATRAVÉS DE DADOS ABERTOS

Engenharia de Dados

**João Augusto Costa Branco Marado Torres**

Trabalho realizado no âmbito da unidade curricular de Sistemas de Informação

ORIENTAÇÃO

Dr.ª Isabel Sofia Sousa Brito

Beja, novembro de 2025

## **Júri**

Responsável: Dr.<sup>a</sup> Isabel Sofia Sousa Brito

Vogal: Dr. João Paulo Trindade

Vogal: Dr.<sup>a</sup> Elsa da Piedade Chinita Soares Rodrigues

# Conteúdo

Conteúdo	i
1 Introdução	1
2 Desenvolvimento	4
3 Conclusão	5
Licença	6

# 1 Introdução

Eu uso Linux e não Microsoft Windows já faz 5 anos. Por várias razões, mas tudo tem a mesma raiz do problema: Windows é software proprietário.

Eu vou sempre que conseguir usar FLOSS. Muitas vezes tenho que ceder aos monopólios do software, porque não o fazer prejudica-me na vida fora do digital. Por exemplo, parece que muita da infraestrutura informática do politécnico depende da Microsoft, a começar pelos correios eletrónicos disponibilizados para os alunos, eu sendo aluno, tenho que o usar porque senão perco acesso a informação importante para completar a minha formação. E é assim que estes oligarcas digitais criam dependências. Mas eles são capazes de se introduzir ainda mais na vida académica dos estudos. Quer dizer, caraças, eles metem-se em tudo o que conseguem, seja para controlar ou ser controlado, desde que haja lucro sendo feito.

Na pedagogia e ensino, devia ser ilegal o uso de software proprietário, já que este vai contra todos os fundamentos da educação. FLOSS tem que ser a norma. O software que deixa o estudante aprender e estudar o software que usa, executar o software da forma que lhe for mais conveniente, modificar para as suas necessidades, e conseguir partilhar a sabedoria que adquiriu com qualquer um.

Acho que já deu para ter uma ideia do porquê da minha indignação. E para minha surpresa, está cadeira pede para o aluno usar Microsoft Excel, Microsoft Power BI e Microsoft SQL Server. Primeiro que eu nem consigo usar alguns desses softwares no meu portátil sem ter que ter o Windows, e é que nem pensar. E depois que existem alternativas, e boas (senão melhores) que são FLOSS. Vou mencionar OpenRefine e Apache Superset que fazem tudo o que era necessário. Não os usei porque, igual as ferramentas da Microsoft, esses não oferecem uma forma simples de reproduzir as mesmas ações em qualquer outra máquina.

Falando mais sobre a cadeira, uns tópicos que eu acho que deviam ser pelo menos mencionados é o dos dados abertos mencionando o esquema da implementação das 5 estrelas e os princípios FAIR (Findable, Accessible, Interoperable, Reusable). Mas claramente não querem saber disso. Incentivam a usar o formato proprietário do Excel, duas estrelas. Não querem nos ensinar a explorar, aprender, adquirir um pensamento crítico independente mas sim a consumir a droga do software proprietário, e aceitar que tem que ser assim, que estas são as ferramentas corretas.

Não querem ensinar como aprender mas sim o que aprender. A resistência em adotar a educação aberta decorre do desejo de controlar o que aprendemos. E nem tentam esconder mais isto. É triste perguntar a razão de estarmos a aprender algo e a resposta ser “porque é o que as empresas querem”.

A minha tarefa inicial é pegar em várias séries temporais, e aplicando o padrão ETL, acabar com um mo-

delo de dados “cubo” (uma base de dados multidimensional em conceito) que permita o modelo de análise OLAP. No final, fazer uma análise da informação possível de obter através do data warehouse.

O tema que escolhi para o trabalho é relacionado à exploração global do trabalho e a desigualdade num nível global, usando parcelas salariais do PIB, coeficiente de Gini... e era interessante conseguir analisar como a divisão do trabalho na dinâmica imperialista e neoliberal atual se manifestam nas acumulações desiguais, dependências imperialistas, precariedade, salários, condições de trabalho...

A minha ideia é provar que existe uma dialética entre a produtividade e parcelas salariais, os trabalhadores trabalhando melhor receberão pior. Tentar relacionar a dependência em trocas comerciais e o IED com a repartição do valor agregado, mostrar como alguns países sugam os recursos de outros. Relacionar a participação nos lucros e resultados com a repartição do valor agregado para calcular o quão explorados estamos a ser numa determinada região do mundo. Encontrar quais são as principais razões para as greves, por exemplo, se o salário não sobe por muito tempo. É apresentar as contradições do capitalismo usando dados abertos e ferramentas FLOSS demonstrando a viabilidade técnica e ética dos mesmos.

Durante a minha pesquisa achei multipas potenciais fontes de dados na Internet capazes de me ajudar nesta tarefa:

**Organização Internacional do Trabalho** Disponibiliza dados sobre empregabilidade, salários, parcelas salariais, sindicalização, horas de trabalho;

**Banco Mundial** Para dados como o PIB, Gini, abertura comercial, fluxos de capital;

**Our World In Data** Oferece diversas informações como a desigualdade, salário mínimo, ou sobre o trabalho informal e desemprego;

**Wikidata** Para qualquer metadado que eu necessite, sobre países, empresas, sindicatos e protestos relacionados ao trabalho;

**Confederação Sindical Internacional** Eles fazem os “ITUC Global Rights Index” cada ano onde eu consigo encontrar informações como violações ao direito dos trabalhadores;

**Base de dados sobre Desigualdade Mundial** Uma base de dados sobre a evolução histórica da distribuição mundial do rendimento e da riqueza, dentro como entre países.

Para o “cubo”, usaria o tempo (anos e décadas) como uma das dimensões, e a geografia (países e regiões, norte e sul global, classificações dos países de acordo com a FMI e a ONU) como outro. Depois talvez também dê para usar os setores de trabalho e se o trabalho é formal ou informal como outra forma de “divisão” de dados para análise. Depois claro informações sobre o trabalho (salários, horas de trabalho, parcelas salariais) e sobre o capital (PIB per capita, fluxos IED, participação nos lucros e resultados).

No futuro tenho que pensar melhor na elaboração de uma tabela de fatos.

Como queria fazer algo reproduzível, decidi usar R ou Python para o ETL, com ajuda de scripts POSIX e Makefiles. Para o load eu achei a base de dados DuckDB, que depois vim a descobrir que basicamente consigo

fazer o ETL inteiro com essa base de dados. Para a parte de análise, Jupyter Notebooks ou R Markdown, ou apenas Python ou R com bibliotecas como Gnuplot ou parecido. Depois consigo gerar PDFs e PostScripts.

No final, se tiver tempo, experimentar codificar os dados para um formato de 4 estrelas no mínimo, RDF ou JSON-LD.

Penso em usar os serviços da Wikimedia Foundation também, caso precise de informação extra, ou se eu quiser contribuir para o projeto com a informação que conseguir retirar da análise.

Durante a análise, vou experimentar várias formas de mostrar as informações: tabelas a partir do *pivot*, correlações, regreções...

Quero depois disponibilizar os dados tratados em formatos abertos, usando URIs sempre que possível para identificadores de países ou anos. Garantir que está tudo FAIR. E documentar neste relatório o processo ETL.

## **2 Desenvolvimento**



## 3 Conclusão

# Licença

Este documento está licenciado sob uma Licença Creative Commons Atribuição–Partilha nos Mesmos Termos 4.0 Internacional (CC BY-SA 4.0).

O código fonte (ficheiros `.tex`, `.bib`, `Makefile`, etc.) utilizado para produzir este relatório está licenciado sob a GNU Affero General Public License v3.0 (AGPL v3).