



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, dezembro 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.<sup>a</sup> Isabel Sofia Sousa Brito

Beja, dezembro 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

<b>Conteúdo</b>	<b>i</b>
<b>Lista de Abreviaturas e Siglas</b>	<b>ii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Objetivos . . . . .	1
1.2 Abordagem e Estrutura do relatório . . . . .	2
<b>2 Desenvolvimento</b>	<b>3</b>
2.1 Enquadramento teórico . . . . .	3
2.2 Metodologia . . . . .	3
2.3 Análise . . . . .	9
2.4 Discussão . . . . .	9
<b>3 Conclusão</b>	<b>11</b>
<b>Referências Bibliográficas</b>	<b>12</b>
<b>Licença</b>	<b>13</b>
<b>A Indicadores</b>	<b>14</b>
A.1 PIB . . . . .	14
A.2 Participação do rendimento do trabalho . . . . .	14
A.3 Índice de Gini . . . . .	14
A.4 Investimento direto estrangeiro . . . . .	14
A.5 Profit sharing . . . . .	15
A.6 Real wages . . . . .	15
<b>B Justificação das escolhas tecnológicas</b>	<b>16</b>
B.1 <i>Software</i> livre . . . . .	16
B.2 Dados Abertos . . . . .	16

# **Lista de Abreviaturas e Siglas**

**ETL** *Extract, Transform, Load*

**FAIR** *Findable, Accessible, Interoperable, Reusable*

**FLOSS** *Free Libre and Open Source Software*

**FMI** Fundo Monetário Internacional

**GUI** *Graphical User Interface*

**IED** Investimento Estrangeiro Direto

**ILO** *International Labour Organization*

**OLAP** *Online Analytical Processing*

**ONU** Organização das Nações Unidas

**PIB** Produto Interno Bruto

**POSIX** *Portable Operating System Interface*

**PPC** Paridade do Poder de Compra

**URI** *Uniform Resource Identifier*

# 1 Introdução

As transformações recentes do capitalismo por conta da crise da ordem têm sido acompanhadas por um agravamento das desigualdades económicas e sociais, particularmente no domínio do trabalho e da distribuição do rendimento à escala mundial, levando à austeridade e reformas laborais. Diferenças persistentes nos níveis salariais e na estabilidade do emprego evidenciam assimetrias estruturais entre países e regiões, onde pessoas vivem bem numa parte do mundo em troca de outras viverem na miséria.

Este é o tema do trabalho: exploração global do trabalho e a desigualdade num nível global.

A disponibilização de *datasets* por múltiplos organismos internacionais e instituições públicas cria possibilidades para a análise destas desigualdades. Em particular, dados ao longo do tempo (*time series*), usando engenharia de dados e análise, permitem observar a evolução histórica de indicadores laborais e identificar tendências de longo prazo. Exemplos de tais indicadores podem ser as participações do rendimento do trabalho no PIB Produto Interno Bruto (PIB), coeficiente de Gini... (ver Apêndice A para definições e fórmulas). Era interessante conseguir analisar como a divisão do trabalho na dinâmica imperialista e neoliberal atual se manifestam nas acumulações desiguais, nas dependências imperialistas, na precariedade, nos salários, nas condições de trabalho...

Apesar da abundância de dados estatísticos disponíveis, a análise da desigualdade laboral global permanece fragmentada. Os dados encontram-se dispersos por múltiplas fontes e são frequentemente analisados de forma isolada, através de indicadores únicos.

Esta separação dificulta a compreensão das relações entre salários, rendimento, condições de trabalho e ação direta laboral (como as greves), bem como a identificação de contradições estruturais do capitalismo.

## 1.1 Objetivos

O objetivo geral deste projeto é investigar a relação entre a produtividade e parcelas salariais. Avaliar a dependência das trocas comerciais e do Investimento Estrangeiro Direto (IED) no *labor share*. Analisar as principais razões para as greves, por exemplo, se o salário não sobe por muito tempo. É apresentar as contradições do capitalismo usando dados *timeseries* abertos num formato multidimensional e ferramentas livres demonstrando a viabilidade técnica e ética dos mesmos. Pretende-se construir um modelo de dados que permita observar, de forma conjunta, diferentes indicadores relacionados com salários, rendimento e condições laborais, entre países ou regiões e ao longo do tempo.

De forma mais específica, o projeto visa:

- recolher e integrar dados provenientes de múltiplas fontes abertas;
- estruturar esses dados num modelo multidimensional adequado à análise;
- produzir análises e visualizações que evidenciem padrões, assimetrias e tendências relevantes.

### 1.1.1 Questões de investigação

Existe uma relação sistemática entre produtividade do trabalho e *labor share*?

Essa relação difere entre países centrais e periféricos?

Indicadores de dependência externa (IED) estão associados a menor *labor share*?

Existe associação entre precariedade e ação direta?

#### 1.1.1.1 Hipóteses

Em contextos de aumento da produtividade do trabalho, a parcela salarial do PIB tende a permanecer estagnada ou a diminuir.

A inserção subordinada na economia global está associada à transferência sistemática de valor do trabalho de países periféricos para países centrais.

## 1.2 Abordagem e Estrutura do relatório

Para alcançar estes objetivos, o trabalho adota uma abordagem baseada em engenharia de dados e o modelo de análise *Online Analytical Processing* (OLAP). O processo inclui a aplicação do padrão *Extract, Transform, Load* (ETL), a construção de um *data warehouse* e a realização de análises multidimensionais via "cubos" (uma base de dados multidimensional em conceito), tabelas e visualizações.

O relatório encontra-se organizado em três capítulos principais. Após esta introdução, o capítulo de desenvolvimento apresenta o enquadramento teórico, a metodologia adotada, a análise dos dados e a discussão dos resultados. Por fim, o capítulo de conclusão sintetiza os principais resultados, discute as limitações do estudo e aponta possíveis direções para trabalhos futuros.

# 2 Desenvolvimento

## 2.1 Enquadramento teórico

Enquadramento metodológico geral

### 2.1.1 Economia política

### 2.1.2 Desigualdade global

### 2.1.3 Divisão internacional do trabalho

### 2.1.4 Limitações dos dados estatísticos

Com informações como estas, como se mede exatamente a exploração, a dependência do imperialismo, a precariedade, ou a extração de valor. Não é bem o objetivo medir a exploração diretamente de dados monetários como o PIB e os outros. Talvez isso nem seja possível e não é só uma questão de eu não saber fazer. O meu objetivo é pegar nesses dados e usá-los para fazer uma análise empírica, quer dizer, observando os dados, e olhando para o mundo em que vivemos, e chegar à conclusão que a minha análise bate certo com a realidade, e é consistente com teorias da economia política. Eu penso que os dados que escolho usar têm uma relação suficientemente forte para conseguir testar a minhas hipóteses, apesar das limitações. Seria parecido às formas de como é medida a qualidade de vida, que não é algo que se meça diretamente.

Outra coisa é que pelos dados serem de fontes diferentes, podem usar definições e metodologias diferentes para calcular o mesmo (teoricamente), que depois pode causar-me problemas se eu não tiver cuidado. É algo com que preciso de ter atenção na etapa 2.2.4.2 do ETL.

## 2.2 Metodologia

### 2.2.1 Escolha das tecnologias e *software*, e técnicas de análise

O meu foco é em criar algo que é facilmente reproduzível, sem ter que te dar indicações de onde tens que clicar na interface gráfica para elaborar uma ação específica. Por isso, decidi usar R ou Python para o ETL, com ajuda de *scripts Portable Operating System Interface* (POSIX) e *Makefiles*.

No início, ainda pensei em usar um simples *script* POSIX. Mas depois achei que ia se transformar em algo muito grande, complicado, e com código repetido. Passei logo para a ideia de usar Makefiles. Eles vão-me permitir orquestrar todos os passos a executar neste trabalho, e caso eu mude algo no *transform*, por exemplo, consigo pedir para executar tudo, mas na realidade ele vai dar passar o passo da extração já que ele não foi afetado por nenhuma mudança, então poupo um pouco de tempo.

Para o *load*, usamos a base de dados DuckDB, que na realidade, ela permite fazer o ETL inteiro através do que ela oferece fora da caixa, e se necessário, extensões. Por exemplo, conseguimos criar tabelas a partir de documentos usando o protocolo HTTP, por exemplo, usar o *statement* UNPIVOT se necessário, e no final, usar a operação de conjuntos `UNION [ALL] BY NAME` numa *query*. Mas nós apenas a vamos usar pelo facto de ela ter sido feita para conseguir fazer operações OLAP de raiz, rapidamente. Com certeza uma experiência para os tempos livres.

---

```

1 CREATE OR REPLACE TABLE flights AS
2   FROM 'https://duckdb.org/data/flights.csv' ;
3
4 UNPIVOT monthly_sales
5 ON COLUMNS(* EXCLUDE (empid, dept))
6 INTO
7   NAME month
8   VALUE sales;
9
10 SELECT * FROM capitals
11 UNION BY NAME
12 SELECT * FROM weather;
```

---

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, regressões...

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

Estas decisões não foram feitas ao calhas. Eu tenho uma ideia muito forte de como todos devemos usar e criar software, incluindo, mas especialmente para trabalhos académicos. A explicação para isso está no Apêndice B.

### 2.2.2 Estrutura do projeto

Terá um ficheiro README.md que servirá como uma introdução para quem pela primeira vez olha para este projeto, sem nenhum contexto.

Na pasta `data` vi estar guardado tudo o que são dados. Teria uma pasta `raw` com os dados acabados de serem retirados da `source` e uma `cleaned` que guardasse as transformações dos anteriores com todas as mudanças necessárias para o seu uso na análise.

O que vai fazer as tarefas mencionadas em cima vão ser uma série de `scripts` localizados na pasta `cleaned` com executáveis.

O código para a criação do relatório e quase tudo relacionado à documentação estaria na pasta `docs`.

Como a minha ideia é usar Python ou R, faz sentido que eu acabe por usar Jupyter Notebooks ou R Markdown para complementar a documentação, isso estará na pasta `notebooks`.

Código SQL fica na pasta `sql`.

### 2.2.3 Fontes de dados

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

**Organização Internacional do Trabalho** Disponibiliza dados sobre empregabilidade, salários, *labor share*, 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 metadados 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.

Dado o caráter dos dados das fontes utilizadas, sem nenhuma relação inicial, a análise baseia-se em indicadores indiretos, entendidos como aproximações empíricas a processos sociais mais complexos, ou seja, vamos relacionar esses valores para termos uma aproximação dos valores que realmente queremos, e que na realidade, são impossíveis de medir: exploração. Como foi falado na 2.1.4

Por exemplo, o PIB representa o valor de riqueza produzido num espaço de tempo num determinado local. Com dados sobre o número de horas trabalhadas durante esse espaço de tempo, eu consigo ter a razão do PIB por hora de trabalho. Quanto maior esse valor, então mais riqueza foi gerada em uma hora, maior a produtividade. Depois temos dados como o *labor share*: qual é a percentagem do rendimento gerado num espaço de tempo num determinado local — pode ser a percentagem do PIB — que foi para quem gerou o rendimento (trabalhadores), por exemplo, em salários, ao invés de ir para o capital. Com isto, posso fazer uma hipótese empírica de que, em economias periféricas, aumentos de produtividade não se traduzem em aumentos proporcionais da parcela salarial.

## 2.2.4 Extract, Transform, and Load

### 2.2.4.1 Extract

Como pôde ser visto em 2.2.3, eu mencionei várias fontes de dados, mas pensei que usar todas elas de uma só vez iria complicar-me a vida muito. Então, para começar, escolhi usar apenas duas: a Organização Internacional do Trabalho — vou começar a mencionar por International Labour Organization (ILO) — e o Banco Mundial para começar. A extração tem que ser feita de modo que no futuro seja simples adicionar outras fontes de dados.

Para o ILO, decidi extrair informações como o *labour income share as a percent of GDP (%)*, a quantidade de contratos de emprego (formais e informais) divididos por sexo, atividade económica e setor (público ou privado) do trabalhador, e a média do salário por hora também dividido por sexo e atividade económica do trabalhador. Do Banco Mundial, de cada país, extraí o PIB (no valor atual do dólar dos Estados Unidos da América), o *Gini index*, e as entradas líquidas de IED (registado na balança de pagamentos e expresso em dólares).

Comecei pelo ILO. Descobri que eles oferecem uma biblioteca R — (Bescond et al., 2025) — para facilitar o meu trabalho. Ficou decidido logo aqui que eu iria usar R ao invés de Python (apesar de que eu já estava inclinado para usar o R). Aprendi o básico de R a partir do (R Core Team, 2025). Depois, vi como podia usar a biblioteca que encontrei, e espreitei as bibliotecas listadas nos *imports* e *suggests* desta para saber o que faziam, já que poderiam dar-me jeito — e deram. Qualquer dúvida, recorri ao («Perplexity AI», s.d.).

Quero mencionar que bem no início era para eu usar o (OpenRefine Community, 2025), por ser o que me parecia o mais semelhante ao Power Query. Ele permite o *extract* e *transform* dos dados, e essas ações, ficam guardadas num histórico, num ficheiro JSON, que serviria para reproduzir as operações entre computadores. Mas isso requisita usar a *Graphical User Interface* (GUI) mesmo assim; já não achei reproduzível o suficiente. Entretanto, passei também os olhos na possibilidade de usar (Apache Software Foundation, s.d.),

(Spotify AB, s.d.) e (Prefect Technologies, s.d.). Ficam aqui as menções.

Criei o ficheiro `extract.R` que apenas fazia o *download* dos *datasets*.

Passando para o Banco Mundial, achei que seria mais simples apenas usar (Nikšić, s.d.). Renomeei o ficheiro anterior para `extract_ilostat.R` e criei o `extract_worldbank.sh`. Descobri logo que o *link* que eles disponibilizavam na API para fazer o *download* do *dataset* em *comma separated values* na realidade era um arquivo comprimido, mas lá dei a volta a esse problema. Com o ficheiro com os dados, vem um com metadados para cada código de país, e um com uma descrição sobre a informação que eu tinha carregado da Internet.

Comecei a criar o `Makefile`. Criei uma regra que executa todos os ficheiros na pasta `bin` que comecem com `extract_`. Para não ter que estar sempre a editar o ficheiro e os `targets` com uma mudança na extração, usei a criação de um ficheiro sem nada escrito, mas que a presença dele indica que o E do ETL já foi concluído, e que depois pode ser usado para dependências.

#### 2.2.4.2 *Transform*

Usei o mesmo formato de nomear os ficheiros, e a mesma ideia no `Makefile`, mas agora a regra `transform` depende do `extract`.

Usei algumas das sugestões de bibliotecas que encontrei anteriormente, nomeadamente o (Wickham et al., 2023) para facilitar a manipulação dos *data frames* e (Wickham & Posit Software, 2025) para manipular *strings*. Também encontrei o (Wickham et al., 2025) que me deixa usar colunas dos *data frames* como variáveis, facilitando o processo todo.

Nos dados do ILO queria renomear as colunas em todos os ficheiros para o mesmo, e remover a linha que podia ser calculada a partir das outras (o total do setor económico).

Para o Banco Mundial, quis renomear os ficheiros para algo menor que explicasse o que estava presente neles. Tive que fazer um *unpivot* criando a coluna `year`. Mudei o nome de algumas colunas também para ficar igual ao do ILO e juntei alguns dos metadados que tinham vindo juntos.

Foi inicialmente adotada uma estratégia de consolidação dos indicadores num único conjunto de dados antes do carregamento. Criei um único ficheiro — `integrate.R` — onde eu leio cada ficheiro de dados, e uso o código do país e no ano de cada linha para criar linhas que agregam todos os valores correspondentes a esses dois que os ficheiros separadamente possuíam. No *data frame* do salário por hora, tive que criar 3 novas colunas via *pivot*, um para cada tipo de representação do salário, e depois escolhi usar apenas o método Paridade do Poder de Compra (PPC), que me pareceu a melhor opção comparado com o valor em dólar ou em moeda social.

No entanto, esta abordagem revelou-se inadequada do ponto de vista de modelação multidimensional. Uma tabela não normalizada não é capaz de se tornar numa *data warehouse*. Estaríamos a adicionar uma etapa a mais ao processo ETL que não existem e é conceitualmente errado. Não existe tabelas de factos, nem de dimensões. Não existem chaves primárias nem estrangeiras. E também não existe um *star/snowflake schema*.

Desnormalização e agregação são tarefas que só vão ser necessárias mais tarde durante a análise.

Sem uma *data warehouse*, não temos modelo multidimensional, não conseguimos fazer OLAP, e torna-se complicado fazer análises.

A tarefa aqui no *transform* é preparar os dados para a criação da *data warehouse* na próxima etapa *load*. Então para isso já temos que ter um modelo multidimensional em mente, o qual foi o que faltou anteriormente, e por isso criamos uma etapa que estraga o propósito do trabalho. O nosso modelo multidimensional é detalhadamente explicado na 2.2.5.

Sempre a atualizar o *Makefile*.

Neste momento existiam muitas linhas cheias de **NA**, exatamente porque essa informação não existe. O que fazer? Por serem tudo números, uma das opções preencher todos os espaços vazios com 0. Mas isso não faz sentido.

- PIB 0 significava que nenhum dinheiro foi gerado, impossível;
- Coeficiente Gini 0 era dizer que não existiam desigualdades, impossível no capitalismo;
- Participação do rendimento do trabalho no PIB de 0 é algum tipo de escravidão;
- Média do salário por hora ser 0 também é exploração total do trabalhador.

IED 0 até é possível existir, mas improvável — o fluxo de entrada e saída tinha que ser o mesmo, anulando-se —, e a ausência do valor não significa que ele não existiu, apenas que não foi calculado, ou sei lá.

Devo remover as linhas que têm pelo menos uma falta de valor então? Não, porque isso normalmente acontece nos países do sul global, e isso tiraria a sua representatividade no *dataset*.

Existem muitas formas de tratar os valores em falta. Se falta dados em anos que ficam entre outros anos onde temos dados, podíamos interpolar esses valores com ajuda de *machine learning*. Neste caso, apenas deixamos as linhas da forma que elas estão, sem apagar nada também, por forma a preservar representatividade dos países periféricos. Na altura da análise, é decidido o que fazer em cada ocasião com os dados em falta.

#### 2.2.4.3 *Load*

A decisão ficou entre («DuckDB», 2025) e («TimescaleDB», 2025) para PostgreSQL, que me lembrei de ambas as menções de vídeos separados do canal de YouTube (FireShip, 2025).

Optei pelo DuckDB por ser mais minimalista, estilo SQLite, enquanto para o outro eu ia precisar de um servidor PostgreSQL.

A cena do DuckDB é que foi desenhado para suportar OLAP tratando as tabelas não como um conjunto de linhas, mas mais como um conjunto de colunas, e por isso é rápido nas operações OLAP.

Para esta parte do processo ETL, apenas escrevi SQL que o DuckDB entende, e o *script load.sh* manda executar esse código no DuckDB e cria um ficheiro de base de dados persistente.

## 2.2.5 Modelo multidimensional

O modelo multidimensional proposto organiza os dados em torno de uma tabela de factos centrada em indicadores económicos e laborais, articulada com dimensões temporais, geográficas e estruturais.

### 2.2.5.1 Tabelas de factos e dimensões

A tabela de fatos é composta por indicadores económicos com relação ao trabalho. Teria como medidas o *labor share*, as *real wages*, o Gini, a produtividade, fluxos de IED, e dias de greve. As dimensões seriam o tempo (ano, década), país, região, setor, tipo de trabalho (formal ou informal).

Para o "cubo", usaria essas dimensões, como o tempo e a geografia — países e regiões, norte e sul global, classificações dos países conforme o Fundo Monetário Internacional (FMI) e a Organização das Nações Unidos (ONU) —, como forma de "divisão" de dados para análise.

## 2.3 Análise

Na minha pesquisa para encontrar a melhor substituição ao Power BI, achei o seguinte:

- Metabase;
- Apache Superset;
- Grafana;

Todos depois teriam o mesmo problema que achei com OpenRefine.

Decidi que esta seria a melhor altura para usar RMarkdown e bibliotecas como o `ggplot2` que também tinha sido recomendado. Mas como não tive tempo para aprender como usar RMarkdown, usei apenas um RScript.

No futuro pretendo experimentar o Streamlit.

### 2.3.1 OLAP

### 2.3.2 Resultados principais

## 2.4 Discussão

Na União Europeia, por exemplo, para conseguirem ter mais dinheiro para gastar em fins bélicos, vemos acontecer por vários países da federação, cortes orçamentais na segurança social e na educação e saúde pública, e com isso, reformas nas leis laborais com o intuito de aumentar a produtividade, que em troca, traz o aumento da precariedade do trabalhador e o aumento da desigualdade económica se a parcela salarial não

aumentar com o aumento da produtividade. Por isso vemos na Europa várias greves gerais quase em simultâneo: 2 de dezembro na França, 28 de novembro e 12 de dezembro na Itália, de 24 a 26 de novembro na Bélgica, e aqui em Portugal a 11 de dezembro. Os sindicatos falham ao não relacionarem estes eventos, demonstrando que o problema não desaparece quando passamos as bordas do país, e que a luta e resistência é internacionalista, neste caso, é a nível europeu.

Na União Europeia, por exemplo, para conseguirem ter mais dinheiro para gastar em fins bélicos, vemos acontecer por vários países da federação, cortes orçamentais na segurança social e na educação e saúde pública, e com isso, reformas nas leis laborais com o intuito de aumentar a produtividade, que em troca, traz o aumento da precariedade do trabalhador e o aumento da desigualdade económica se a parcela salarial não aumentar com o aumento da produtividade. Por isso vemos na Europa várias greves gerais quase em simultâneo.

## 3 Conclusão

# Referências Bibliográficas

- Apache Software Foundation. (s.d.). *Apache Airflow*. <https://airflow.apache.org/>
- Bescond, D., Villarreal-Fuentes, M., & of Statistics, I. D. (2025). *Rilostat: R Interface to ILOSTAT Open Data* (Versão 2.3.4). <https://cran.r-project.org/web/packages/Rilostat/index.html>
- DuckDB. (2025). <https://duckdb.org/>
- Fireship. (2025). <https://www.youtube.com/c/Fireship>
- Nikšić, H. (s.d.). *GNU Wget*. <https://www.gnu.org/software/wget/>
- OpenRefine Community. (2025). *OpenRefine*. <https://openrefine.org/>
- Perplexity AI. (s.d.). Perplexity AI, Inc. <https://www.perplexity.ai/>
- Prefect Technologies. (s.d.). *Prefect*. <https://www.prefect.io/>
- R Core Team. (2025). *An Introduction to R*. Versão 4.5.2. R Foundation for Statistical Computing. Vienna, Austria. <https://cran.r-project.org/doc/manuals/r-release/R-intro.html>
- Spotify AB. (s.d.). *Luigi: Workflow Scheduling System*. [https://luigi.readthedocs.io/en/stable/running\\_luigi.html](https://luigi.readthedocs.io/en/stable/running_luigi.html)
- TimescaleDB. (2025). <https://github.com/timescale/timescaledb>
- Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., & Posit Software, P. (2023). *dplyr: A Grammar of Data Manipulation* (Versão 1.1.4). <https://cran.r-project.org/web/packages/dplyr/index.html>
- Wickham, H., & Posit Software, P. (2025). *stringr: Simple, Consistent Wrappers for Common String Operations* [R package version 1.6.0]. <https://cran.r-project.org/web/packages/stringr/index.html>
- Wickham, H., Vaughan, D., Girlich, M., Ushey, K., & Posit Software, P. (2025). *tidyverse*. <https://tidyverse.org/>

# 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).

# A Indicadores

## A.1 PIB

GDP

GDP (current US\$)

Gross domestic product is the total income earned through the production of goods and services in an economic territory during an accounting period. It can be measured in three different ways: using either the expenditure approach, the income approach, or the production approach. This indicator is expressed in current prices, meaning no adjustment has been made to account for price changes over time. This indicator is expressed in United States dollars.

## A.2 Participação do rendimento do trabalho

Labour (income) share

## A.3 Índice de Gini

Gini index

Gini index measures the extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution. A Lorenz curve plots the cumulative percentages of total income received against the cumulative number of recipients, starting with the poorest individual or household. The Gini index measures the area between the Lorenz curve and a hypothetical line of absolute equality, expressed as a percentage of the maximum area under the line. Thus a Gini index of 0 represents perfect equality, while an index of 100 implies perfect inequality.

## A.4 Investimento direto estrangeiro

Investimento direto estrangeiro, entradas líquidas (Balança de Pagamentos, US\$ atuais)

Foreign direct investment, net inflows (BoP, current US\$)

Foreign direct investment refers to direct investment equity flows in the reporting economy. It is the sum of equity capital, reinvestment of earnings, and other capital. Direct investment is a category of cross-border investment associated with a resident in one economy having control or a significant degree of influence on the management of an enterprise that is resident in another economy. Ownership of 10 percent or more of the ordinary shares of voting stock is the criterion for determining the existence of a direct investment relationship. Data are in current U.S. dollars.

## A.5 Profit sharing

## A.6 Real wages

# B Justificação das escolhas tecnológicas

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.

## B.1 *Software* livre

Eu vou, sempre que conseguir, usar *free, libre, and open source software* (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 à informação importante para completar a minha formação. E é assim que estes oligarcas digitais criam dependências nos seus *software*. Mas eles são capazes de se introduzir em muito mais do que apenas na vida académica e nos estudos. Quer dizer, caraças, eles metem-se em tudo o que conseguem, 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. *Software* livre 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 dei para ter uma ideia do porquê da minha indignação. E para minha surpresa, esta 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 como sistema operativo — isso é que nem pensar. E depois que existem alternativas, e boas (senão melhores) que são livres. Vou mencionar OpenRefine e Apache Superset que fazem tudo o que era necessário. Não os usei porque, igual às ferramentas da Microsoft, esses não oferecem uma forma simples de reproduzir as mesmas ações em qualquer outra máquina.

## B.2 Dados Abertos

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 *Findable, Accessible,*

*Interoperable, Reusable* (FAIR). 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”.