



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.^a Isabel Sofia Sousa Brito

Beja, dezembro de 2025

Júri

Responsável: Dr.^a Isabel Sofia Sousa Brito

Vogal: Dr. João Paulo Trindade

Vogal: Dr.^a Elsa da Piedade Chinita Soares Rodrigues

Conteúdo

Conteúdo	i
Lista de Abreviaturas e Siglas	ii
1 Introdução	1
1.1 Objetivos	1
1.2 Abordagem e Estrutura do trabalho	2
2 Desenvolvimento	3
2.1 Enquadramento teórico	3
2.2 Metodologia	3
2.3 Análise	5
2.4 Discussão	5
3 Conclusão	7
Licença	8
A Indicadores	9
A.1 PIB	9
A.2 Participação do rendimento do trabalho	9
A.3 Índice de Gini	9
A.4 Investimento direto estrangeiro	9
A.5 Profit sharing	10
A.6 Real wages	10
B Justificação das escolhas tecnológicas	11
B.1 <i>Software</i> livre	11
B.2 Dados Abertos	11

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 Labor 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 trabalho

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 trabalho 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

2.1.1 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 o 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 2.2.3.2 do ETL.

2.2 Metodologia

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

Como queria fazer algo reproduzível, decidi usar R ou Python para o ETL, com ajuda de *scripts Portable Operating System Interface* (POSIX) e *Makefiles*. Para o *load* eu achei a base de dados DuckDB, que depois descobri que basicamente consigo fazer o ETL inteiro apenas usando 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, 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 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 carácter 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 2.1.1

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.

2.2.3 *Extract, Transform, and Load*

2.2.3.1 *Extract*

2.2.3.2 *Transform*

2.2.3.3 *Load*

2.2.4 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.

2.2.4.1 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 Unidas (ONU) —, como forma de "divisão" de dados para análise.

2.3 Análise

2.3.1 OLAP

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

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 deu 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”.