

**FUNDAÇÃO GETULIO VARGAS**  
**CIÊNCIA DE DADOS E INTELIGÊNCIA ARTIFICIAL**

Cleomar Felipe Rabelo Antosczyszyn

**DESENVOLVIMENTO DE UMA FERRAMENTA PARA FACILITAR O ACESSO E A  
ANÁLISE DOS MICRODADOS DA PNAD CONTÍNUA**

Rio de Janeiro

2025

**CLEOMAR FELIPE RABELO ANTOSZCZYSZYN**

**DESENVOLVIMENTO DE UMA FERRAMENTA PARA FACILITAR O ACESSO E A  
ANÁLISE DOS MICRODADOS DA PNAD CONTÍNUA**

Trabalho de conclusão de curso apresentado  
como requisito parcial para a obtenção do  
título de Bacharel em Ciência de Dados e  
Inteligência Artificial.

Orientador: Julio Cesar Chaves

Coorientador: Paulo Cezar P. Carvalho

Rio de Janeiro

2025

## **AGRADECIMENTOS**

Gostaria de expressar minha profunda gratidão às pessoas que, com sua expertise e generosidade, contribuíram significativamente para o desenvolvimento deste trabalho.

Agradeço imensamente à Janaina Rodrigues Feijó e à Ana Paula Nothen Ruhe, do IBRE, e ao Rafael Guerreiro Osório, do IPEA, pelas inúmeras reuniões e orientações valiosas ao longo do projeto. O conhecimento e a dedicação de todos foram essenciais para a evolução deste trabalho.

A todos, o meu sincero agradecimento por todo o tempo, paciência e colaboração.

## **RESUMO**

Este trabalho apresenta a biblioteca pnadcontinua, desenvolvida para facilitar o acesso e a análise dos microdados da PNAD Contínua. A ferramenta, implementada em Python com interface gráfica Tkinter, permite selecionar variáveis, aplicar filtros, configurar agregações e exportar resultados em CSV. Os dados são processados no formato Parquet, garantindo eficiência de armazenamento e leitura. Disponível no GitHub e via PyPI, a biblioteca vai de encontro ao anseio de pesquisadores não técnicos que necessitam de dados da PNAD Contínua para embasar pesquisas sócio-econômicas no Brasil.

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>4</b>
<b>2 REVISÃO TEÓRICA.....</b>	<b>4</b>
2.1 A PNAD Contínua.....	4
2.2 Desafios no Uso dos Microdados da PNAD Contínua.....	5
2.3 Trabalhos Relacionados.....	5
<b>3 DESENVOLVIMENTO.....</b>	<b>6</b>
3.1 Descrição do Problema e Limitações Enfrentadas.....	6
3.2 Redefinição do Escopo.....	7
3.3 Desenvolvimento da Ferramenta.....	7
3.3.1 Estrutura da Biblioteca.....	8
3.3.2 Criação da Interface Gráfica.....	8
3.3.3 Utilização de Arquivos Parquet.....	9
<b>4 RESULTADOS.....</b>	<b>9</b>
4.1 Funcionalidades.....	9
4.1.1 Microdados.....	10
4.1.2 Baixar Agregações.....	12
4.1.3 Descrição de Variáveis.....	13
4.4.4 Tutorial.....	14
4.3 Exemplos de Uso.....	15
<b>5 CONCLUSÃO.....</b>	<b>17</b>
<b>6 REFERÊNCIAS.....</b>	<b>19</b>

## 1 INTRODUÇÃO

A Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua) é uma das principais fontes de dados socioeconômicos no Brasil, fornecendo informações detalhadas sobre características da população, mercado de trabalho, renda e outras dimensões sociais. Esses dados, disponibilizados em formato fixed-width file (FWF), representam um enorme potencial para análises, mas também apresentam desafios significativos para manuseio e interpretação devido ao seu grande volume e formato pouco amigável.

Inicialmente, este projeto tinha como objetivo complementar o trabalho do pesquisador Rafael Osório, do Instituto de Pesquisa Econômica Aplicada (Ipea), utilizando o pacote Pynad, desenvolvido em Python, para criar um modelo de machine learning que identificasse longitudinalmente os respondentes da PNAD Contínua ao longo de diferentes visitas. No entanto, durante o desenvolvimento, identificou-se que as características dos dados tornavam inviável a criação de um modelo eficiente para essa finalidade.

Diante dessas limitações, o projeto foi redirecionado para atender a outra necessidade: facilitar o acesso e a análise dos microdados da PNAD Contínua. Assim, foi desenvolvida uma biblioteca em Python que, por meio de uma interface gráfica intuitiva, permite a importação, seleção e armazenamento dos dados de interesse. Para lidar com o grande volume de informações, que inicialmente ocupa cerca de 100 GB no formato FWF, a biblioteca utiliza o formato Parquet, reduzindo significativamente o espaço necessário e otimizando o desempenho no manuseio dos dados. Essa abordagem busca proporcionar uma solução eficiente e prática para pesquisadores que utilizam a PNAD Contínua.

## 2 REVISÃO TEÓRICA

### 2.1 A PNAD Contínua

A Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua) é uma pesquisa do Instituto Brasileiro de Geografia e Estatística (IBGE), que tem como objetivo fornecer informações contínuas e atualizadas sobre as condições de vida da população brasileira, especialmente relacionadas ao mercado de trabalho. Lançada em 2016, a PNAD Contínua substitui a antiga PNAD, que antes era realizada anualmente.

A pesquisa é fundamental para a elaboração de indicadores sociodemográficos, que são utilizados por órgãos governamentais, empresas, pesquisadores e sociedade em geral para planejar políticas públicas e ações voltadas ao bem-estar da população. A PNAD Contínua

coleta dados sobre o mercado de trabalho, como a taxa de desemprego, ocupação, rendimento, entre outros, além de informações sobre educação, migração, e outras características demográficas.

Com a implementação da coleta contínua de dados, a PNAD Contínua permite uma visão mais atualizada e precisa das condições do mercado de trabalho no Brasil, realizando a coleta em todos os trimestres do ano. Essa continuidade na coleta de dados proporciona uma melhor análise das tendências do mercado de trabalho e facilita o acompanhamento de políticas públicas.

## 2.2 Desafios no Uso dos Microdados da PNAD Contínua

O uso dos microdados da PNAD Contínua apresenta diversos desafios, principalmente em relação ao formato dos dados, ao grande volume de informações e à integração dos metadados. O formato FWF (Fixed Width Format), utilizado pelo IBGE na disponibilização dos dados, é um dos principais obstáculos. Esse formato não é tão comum quanto outros, como o CSV, o que faz com que algumas ferramentas de análise de dados não o suportem diretamente. Como resultado, é necessário realizar processos adicionais de conversão ou manipulação dos dados para torná-los utilizáveis em ferramentas mais populares. Além disso, o grande volume de dados disponibilizados, consistindo em dezenas de gigabytes, torna o processamento e a análise desses dados um desafio significativo, sendo que a infraestrutura necessária para armazenar e manipular essas grandes quantidades de informações exigem recursos computacionais robustos. Por fim, a incorporação dos metadados também representa um desafio importante. Embora os metadados sejam essenciais para a correta interpretação dos dados, como o significado dos campos e as unidades de medida, a integração desses metadados de forma eficiente não é tão simples, pois essas informações são reunidas em uma planilha externa.

## 2.3 Trabalhos Relacionados

Um estudo relevante relacionado ao uso de microdados da PNAD Contínua é o trabalho de Rafael Osório, apresentado no documento "Sobre a Montagem de Microdados Longitudinais da PNAD Contínua" (Osório, 2019), publicado pelo Instituto de Pesquisa Econômica Aplicada (IPEA). O estudo de Osório aborda a construção de um modelo de dados longitudinais a partir dos microdados da PNAD Contínua, com o objetivo de acompanhar as

mudanças nas características da população brasileira ao longo do tempo. A pesquisa de Osório destaca a importância da utilização dos microdados para análise de tendências e da mobilidade social, além de apresentar uma metodologia para a construção de séries temporais com base na estrutura transversal dos dados.

Osório discute também as dificuldades encontradas ao lidar com os dados da PNAD Contínua, como a integração das diferentes bases de dados e o tratamento de variáveis repetidas ao longo do tempo. Ele propõe soluções para superar esses obstáculos, com o uso de ferramentas que permitem a reestruturação dos dados em formato longitudinal, facilitando a análise das mudanças individuais ao longo dos períodos.

Esse trabalho é relevante para o desenvolvimento de ferramentas voltadas para o tratamento de microdados da PNAD Contínua, como a proposta deste TCC, uma vez que ambos buscam facilitar o uso e a interpretação dos dados para análises mais complexas. No entanto, a abordagem de Osório foca mais na construção de séries temporais e modelos de dados longitudinais, enquanto o objetivo deste trabalho é simplificar o acesso e análise dos microdados para pesquisadores, por meio de uma ferramenta que lida com as dificuldades práticas do uso dos dados, como o formato FWF e a integração dos metadados.

### **3 DESENVOLVIMENTO**

#### **3.1 Descrição do Problema e Limitações Enfrentadas**

A proposta inicial deste trabalho consistia em desenvolver um modelo de machine learning para identificar respondentes da PNAD Contínua ao longo das diferentes rodadas da pesquisa, utilizando uma abordagem de identificação longitudinal. Essa proposta tinha como base o pacote Pynad, desenvolvido por Rafael Osorio, que fornece ferramentas para o tratamento dos microdados da PNAD Contínua. A ideia era explorar o potencial do aprendizado de máquina para classificar os indivíduos, superando limitações associadas a métodos tradicionais de identificação dos respondentes.

No entanto, durante o desenvolvimento, foram identificados desafios significativos que dificultaram a viabilidade da proposta. O primeiro obstáculo foi a natureza das variáveis disponíveis nos microdados. Variáveis como Unidade Primária de Amostragem (UPA), Unidade da Federação (UF) e número de identificação do domicílio ou indivíduo possuem uma função primária de identificação, mas não apresentam características explicativas diretamente úteis para o aprendizado de máquina. Para contornar essa limitação, seria

necessário criar variáveis derivadas, como verificar se o ano de nascimento de um respondente permanece constante ao longo das pesquisas.

Essa abordagem, contudo, introduz dois problemas importantes. O primeiro é o aumento considerável da dimensionalidade dos dados, pois a criação de variáveis derivadas para capturar possíveis padrões geraria um grande número de colunas, o que poderia levar ao problema de alta dimensionalidade, prejudicando o desempenho e a eficiência do modelo. Além disso, essas variáveis derivadas são essencialmente arbitrárias, pois refletem decisões específicas definidas no âmbito da pesquisa sobre quais padrões devem ser avaliados. Essa arbitrariedade reduz a capacidade do modelo de capturar padrões fora dos limites definidos previamente, comprometendo a generalização e a aplicabilidade das previsões.

### 3.2 Redefinição do Escopo

O projeto foi redefinido para atender a uma nova finalidade: facilitar o acesso às informações contidas nos microdados da PNAD Contínua por meio de uma interface gráfica. Essa interface permite a realização de agregações nos dados de maneira simples e intuitiva, eliminando a necessidade de conhecimentos avançados em programação ou manipulação de dados. O foco principal passou a ser o desenvolvimento de uma ferramenta que simplifique a extração, seleção e visualização das informações mais relevantes, garantindo praticidade e eficiência no trabalho com grandes volumes de dados.

### 3.3 Desenvolvimento da Ferramenta

A ferramenta criada neste trabalho foi desenvolvida como uma biblioteca em Python, linguagem amplamente adotada pela comunidade científica por sua flexibilidade e eficácia na manipulação e análise de dados. Essa escolha se justifica não apenas pela popularidade do Python, mas também por sua vasta gama de bibliotecas complementares que potencializam o desenvolvimento de soluções robustas e eficientes.

A instalação da biblioteca é intuitiva e realizada através do pip, o gerenciador de pacotes padrão do Python. Com apenas um comando, os usuários conseguem integrar a ferramenta ao seu ambiente de trabalho, garantindo praticidade e agilidade na adoção.

### 3.3.1 Estrutura da Biblioteca

A biblioteca desenvolvida é composta por arquivos e diretórios organizados para facilitar a manipulação dos microdados da PNAD Contínua e a configuração da interface gráfica. A seguir, apresenta-se a estrutura e a funcionalidade de cada componente:

- core.py: Este arquivo contém as principais funções relacionadas à requisição, manipulação e organização dos microdados da PNAD Contínua. Ele é o núcleo da biblioteca, responsável por processar os dados e permitir seu uso de maneira simplificada.
- constants.py: Armazena todas as constantes utilizadas no código-fonte, centralizando variáveis que precisam ser acessadas de forma recorrente, como URLs e parâmetros fixos.
- downloaded.json: Arquivo no formato JSON que registra informações sobre os microdados já baixados. Esse recurso é utilizado para evitar a duplicidade de downloads e gerenciar os dados localmente.
- metadata.py: Este módulo organiza os metadados dos microdados da PNAD Contínua, como os nomes das colunas, suas larguras e outros detalhes relevantes para a interpretação dos dados.
- utils.py: Contém funções auxiliares e utilitárias que oferecem suporte às demais funcionalidades da biblioteca, como formatação, validação e outros processos complementares.
- gui/: Diretório que concentra os arquivos responsáveis pela configuração e funcionamento da interface gráfica da biblioteca. Essa interface é a principal forma de interação do usuário com as funcionalidades oferecidas pela ferramenta.

### 3.3.2 Criação da Interface Gráfica

A interface gráfica da ferramenta foi desenvolvida utilizando o framework Tkinter, uma biblioteca nativa do Python para a criação de interfaces gráficas. A interface gráfica foi projetada para facilitar o acesso às funcionalidades da biblioteca, como a escolha de variáveis de interesse, a aplicação de filtros e a visualização das informações disponíveis. Através de botões, caixas de seleção e menus interativos, os usuários conseguem realizar operações complexas de maneira visual e acessível.

### 3.3.3 Utilização de Arquivos Parquet

Para otimizar o armazenamento e o processamento dos microdados da PNAD Contínua, a biblioteca converte os arquivos no formato FWF (Fixed Width Format) para o formato Parquet. Esse processo reduz significativamente o espaço em disco utilizado e torna mais eficiente a manipulação dos dados.

O formato Parquet organiza os dados de maneira colunar, permitindo que apenas as colunas necessárias sejam carregadas durante a análise. Isso resulta em maior eficiência tanto no uso da memória quanto no desempenho de leitura e escrita, especialmente em bases de dados com muitas variáveis.

Para a leitura e manipulação dos arquivos Parquet, a biblioteca utiliza o Polars, uma ferramenta de processamento de dados em alta performance escrita em Rust e integrada ao Python. O Polars é otimizado para lidar com grandes volumes de dados, oferecendo funcionalidades rápidas e eficientes, como filtragem, agregação e transformação. Essa integração entre Parquet e Polars proporciona uma experiência fluida e ágil, permitindo que os usuários processem os microdados de maneira eficaz, mesmo em máquinas com recursos limitados.

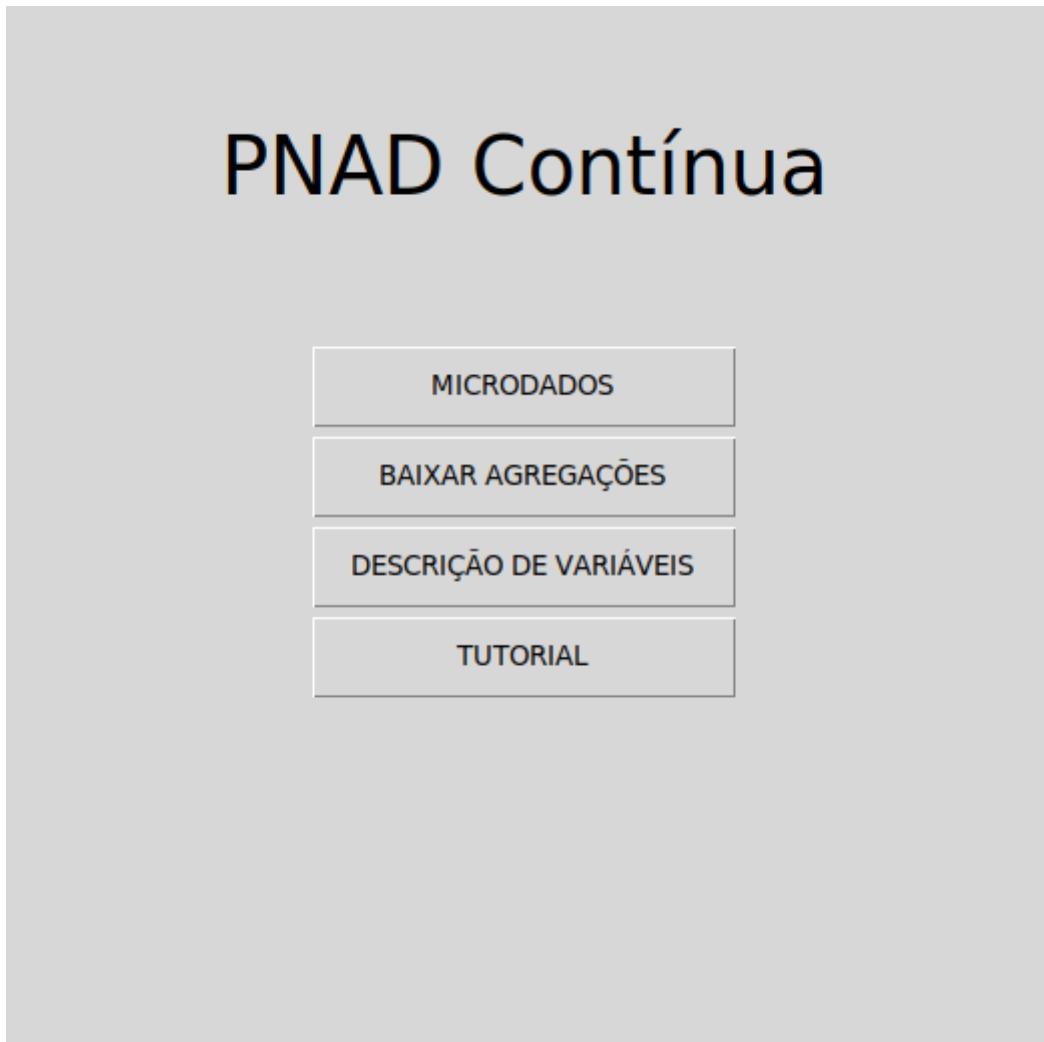
## 4 RESULTADOS

A biblioteca criada, nomeada *pnadcontinua*, está disponível no GitHub, no endereço "<https://github.com/cleomarfelipe/pnadcontinua/>". Os usuários podem acessar o código-fonte, fazer contribuições ou consultar a documentação diretamente nesse repositório.

Além disso, a biblioteca pode ser facilmente instalada via pip com o comando "*pip install pnadcontinua*".

### 4.1 Funcionalidades

A tela inicial da biblioteca apresenta quatro principais seções, acessíveis de forma simples através da interface gráfica. São elas: Microdados, Baixar Agregações, Descrição de Variáveis e Tutorial. A imagem abaixo ilustra essa página:



**Figura 1** - Tela Inicial da Aplicação

#### 4.1.1 Microdados

A seção Microdados permite ao usuário selecionar e baixar os dados necessários para análise, mas é importante destacar que o acesso aos microdados brutos é restrito. Os microdados baixados são processados e utilizados exclusivamente pelo programa, sendo que o usuário não tem acesso direto aos arquivos brutos da PNAD Contínua.

No primeiro acesso, o usuário deve selecionar as variáveis e os trimestres de interesse para análise. As variáveis podem ser escolhidas entre categorias como "Tudo", "Principais", "Trabalho" e "Educação", facilitando a seleção conforme a área de estudo. Além disso, é possível selecionar os trimestres relevantes para a análise.

Após a seleção, o usuário pode baixar os microdados, com o processo inicial podendo levar cerca de uma hora e ocupar até 1 GB de espaço, dependendo da quantidade de variáveis e

trimestres escolhidos. Em acessos subsequentes, o processo é mais rápido, pois apenas os trimestres atualizados precisam ser baixados.

A interface também oferece funcionalidades para gerenciar os dados baixados, permitindo que o usuário visualize informações sobre os dados armazenados, como o total de espaço utilizado, e apague os dados se necessário. Essa opção de remoção de dados garante que os usuários possam controlar o armazenamento local, com a ressalva de que, ao apagar os dados, será necessário baixá-los novamente.

Essa seção foi projetada para facilitar o acesso às informações necessárias para as análises, garantindo que o processo de obtenção dos microdados seja simples e controlado. A figura abaixo mostra essa tela:

The screenshot displays the 'MICRODADOS' interface. At the top left is a 'Voltar' (Back) button. The main title 'MICRODADOS' is centered at the top. Below it, there are two sections for selecting variables and trimesters, each with a 'Seleção:' label and a 'Tudo' button. The first section shows 'Seleção: 4' and the second shows 'Seleção: 0'. A large 'Baixar' (Download) button is located below these sections. At the bottom, a 'Dados Baixados' (Downloaded Data) section contains 'Variáveis' and 'Trimestres' buttons. The text 'Uso de Armazenamento: 881.75 MB' indicates the current storage usage. An 'Apagar' (Delete) button is located at the bottom right of this section.

**Figura 2 - Tela Microdados**

#### 4.1.2 Baixar Agregações

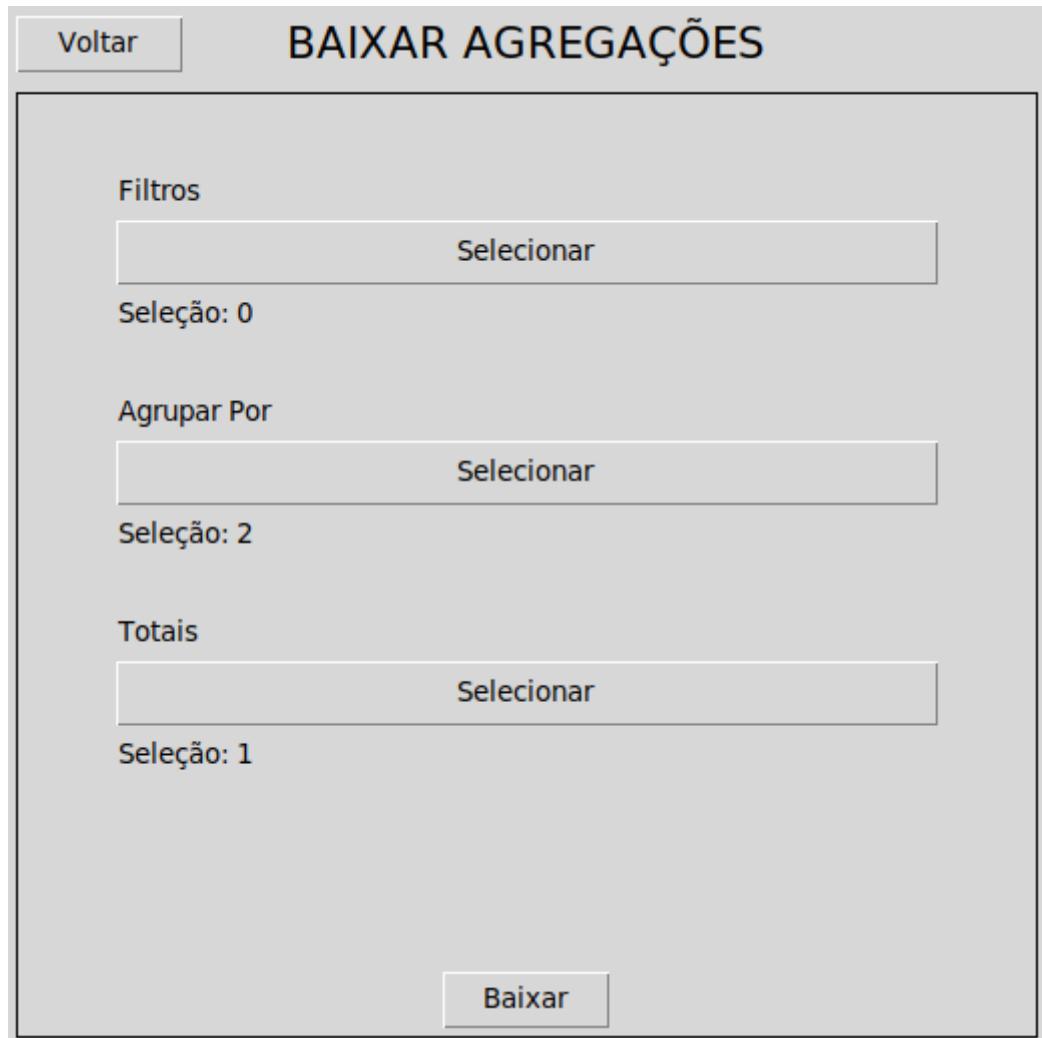
A seção Baixar Agregações permite ao usuário configurar e realizar agregações nos microdados baixados, gerando um arquivo CSV com as informações sumarizadas. Para isso, a página disponibiliza três funcionalidades principais: Selecionar Filtros, Selecionar Grupos e Selecionar Totais.

Primeiramente, ao clicar em Selecionar Filtros, o usuário é direcionado a uma janela com duas abas: Variáveis Categóricas e Variáveis Numéricas. Na aba de Variáveis Categóricas, o usuário pode selecionar as variáveis e categorias que deseja filtrar utilizando checkboxes. Já na aba de Variáveis Numéricas, é possível aplicar um filtro com base em operações numéricas, como maior que, menor que, entre outras, além de definir um valor para o filtro. Para valores decimais, deve-se utilizar ponto (exemplo: 10.5). Após configurar os filtros, o usuário pode adicionar novos filtros clicando em Adicionar ou removê-los clicando no ícone X ao lado do filtro desejado. As escolhas feitas são salvas ao clicar em Voltar.

Em seguida, ao clicar em Selecionar Grupos, o usuário pode escolher as variáveis pelas quais deseja agrupar os dados. Por padrão, as variáveis Ano e Trimestre já estão selecionadas, mas o usuário tem a liberdade de desmarcar essas opções caso não queira agrupar os dados por essas variáveis.

A terceira funcionalidade, Selecionar Totais, permite ao usuário configurar os totais que serão calculados nas agregações. Ao clicar em Selecionar Totais, uma nova janela é aberta com duas abas. Na aba Contagem, o usuário pode optar por incluir ou não a quantidade de pessoas em cada grupo selecionado. Já na aba Soma e Média, é possível calcular a soma ou média de variáveis numéricas para cada grupo. Para variáveis de rendimento Habitual ou Efetivo, uma opção adicional é apresentada para realizar o deflacionamento desses valores antes da agregação, a nível de indivíduo. Após configurar os totais, o usuário clica em Voltar para salvar as escolhas.

Após configurar os filtros, grupos e totais, o usuário pode clicar no botão Baixar para gerar o arquivo CSV com as agregações. No arquivo gerado, as colunas com o sufixo "def" indicam que a variável foi deflacionada, "media" indica que foi calculada a média da variável e "soma" indica que a soma foi calculada. Esse arquivo pode ser utilizado para análise posterior, com as agregações definidas de acordo com as escolhas feitas pelo usuário.



**Figura 3** - Tela Baixar Agregações

#### 4.1.3 Descrição de Variáveis

A seção Descrição de Variáveis permite ao usuário visualizar a descrição de cada variável presente nos microdados, identificadas por seus respectivos códigos. Essa funcionalidade é especialmente útil para entender o significado de cada variável e sua aplicação nas análises realizadas. A imagem abaixo ilustra essa seção.

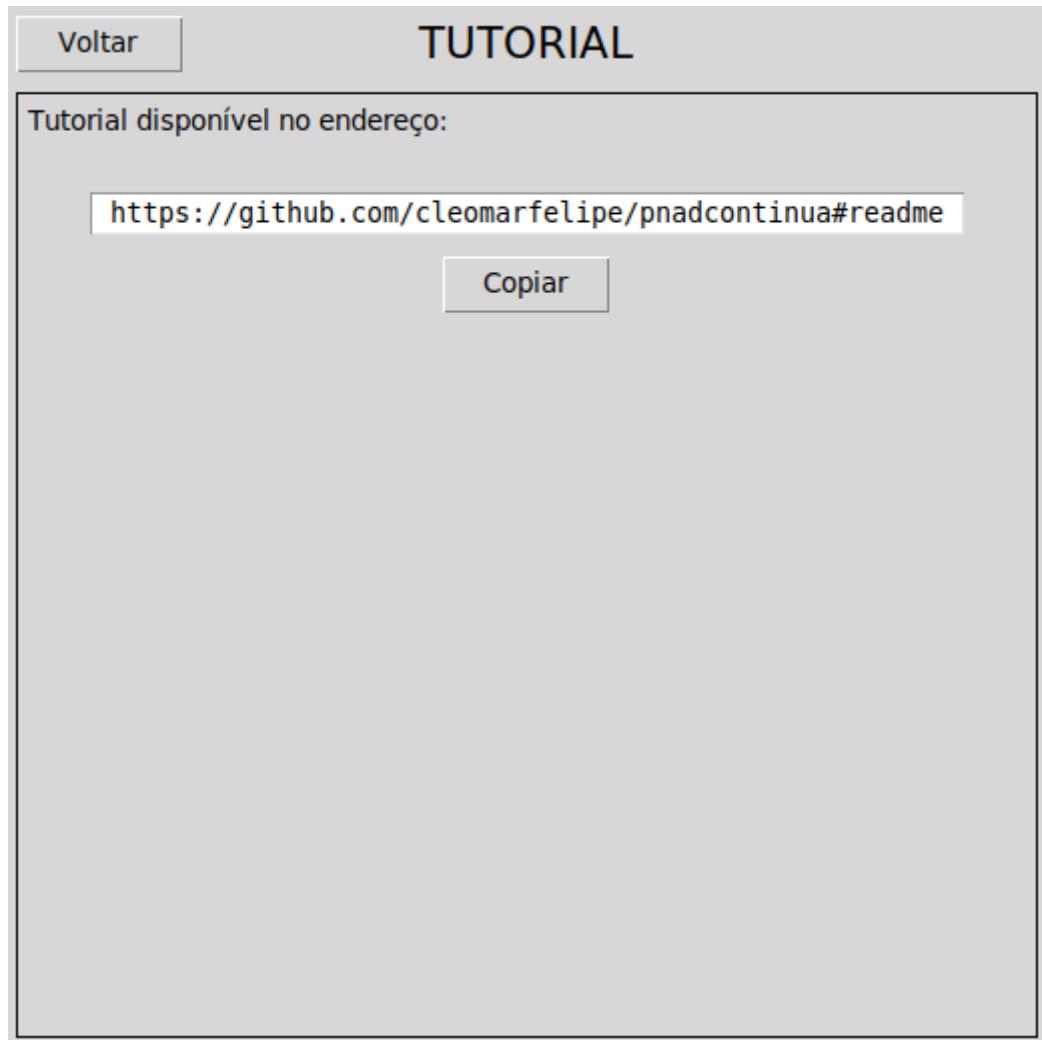
Voltar DESCRIÇÃO DE VARIÁVEIS

- [Ano] Ano de referência
- [Trimestre] Trimestre de referência
- [UF] Unidade da Federação
- [V1028] Peso do domicílio e das pessoas
- [Capital] Município da Capital
- [RM\_RIDE] Região Metropolitana e Região Administrativa Integrada de Desenvolvimento
- [V1022] Situação do domicílio
- [V1023] Tipo de área
- [V2001] Número de pessoas no domicílio
- [V2005] Condição no domicílio
- [V2007] Sexo
- [V2008] Dia de nascimento
- [V20081] Mês de nascimento
- [V20082] Ano de nascimento
- [V2009] Idade do morador na data de referência
- [V2010] Cor ou raça
- [V3001] ... sabe ler e escrever?
- [V3002] ... frequenta escola?
- [V3002A] A escola que ... frequenta é de

**Figura 4** - Tela Descrição de Variáveis

#### 4.4.4 Tutorial

A seção Tutorial fornece um link direto para o repositório oficial da biblioteca no GitHub, onde está disponível um tutorial detalhado sobre como utilizar a ferramenta. Esse guia explica todas as etapas necessárias para aproveitar as funcionalidades oferecidas, desde a importação dos microdados até a seleção e exportação de agregações.

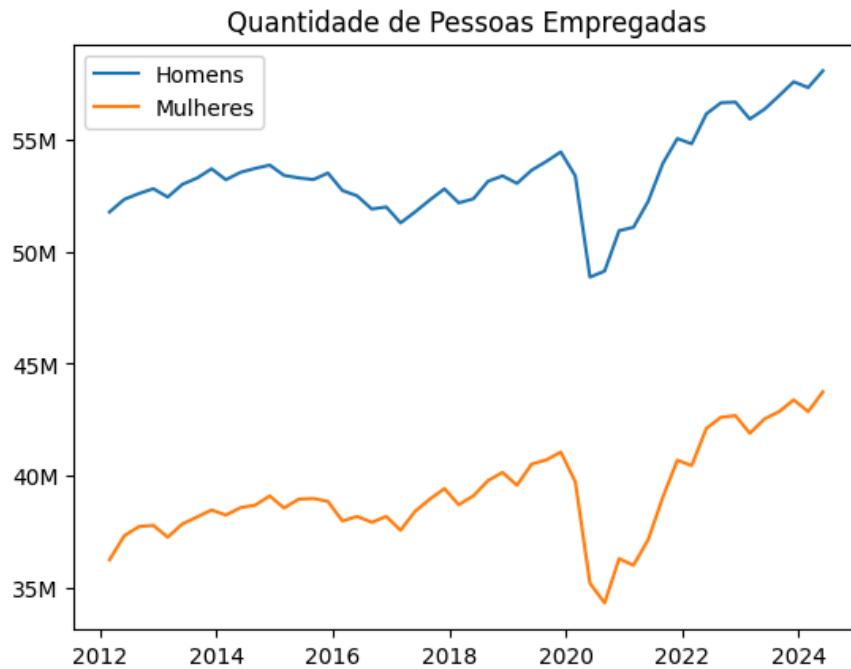


**Figura 5** - Tela Tutorial

#### 4.3 Exemplos de Uso

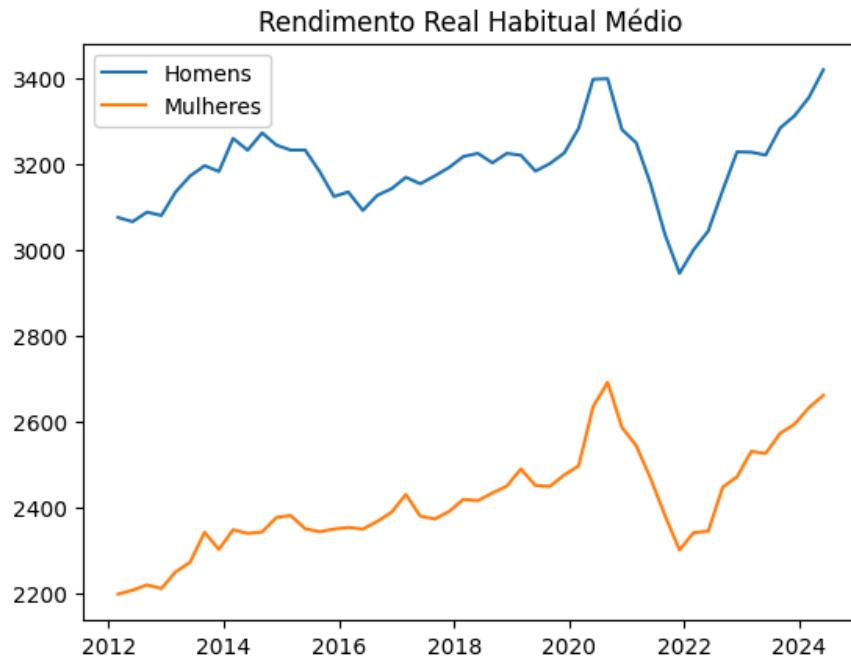
No primeiro exemplo, busca-se calcular o rendimento médio e a quantidade de pessoas ocupadas, segmentadas por ano, trimestre e sexo. Para isso, na interface da ferramenta, o filtro categórico foi aplicado selecionando a variável VD4002 e marcando apenas o valor referente a pessoas ocupadas. Em seguida, foram definidos os grupos para o agrupamento dos dados, selecionando as variáveis Ano, Trimestre e Sexo (V2007). Na configuração dos totais, optou-se por incluir a contagem de pessoas em cada grupo, além de calcular a média da variável VD4016 (rendimento) em duas condições: com e sem deflacionamento. Após configurar os parâmetros, foi gerado um arquivo CSV contendo as colunas Ano, Trimestre, Sexo, a média do rendimento sem deflacionamento (VD4016\_media), a média do rendimento com deflacionamento (VD4016\_def\_media) e a quantidade de pessoas em cada grupo.

(qnt\_pessoas). Com esses dados, podemos construir uma visualização que mostra, por exemplo, a quantidade de homens e mulheres empregados ao longo do tempo:



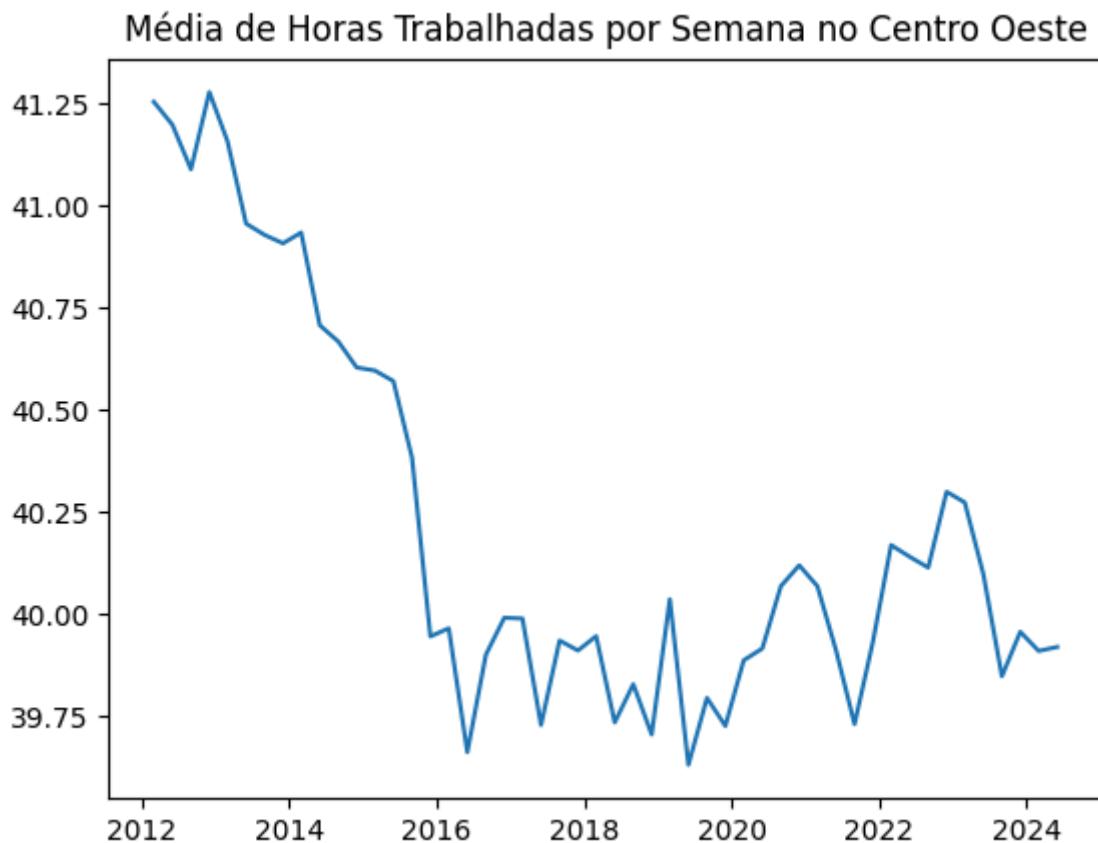
**Figura 6** - Quantidade de pessoas empregadas

E podemos também criar um gráfico que mostra o rendimento real habitual médio, separando por sexo:



**Figura 7** - Rendimento real habitual médio

No segundo exemplo, o objetivo foi calcular a média de horas trabalhadas por pessoas residentes na região Centro-Oeste do Brasil, segmentadas por ano e trimestre. Inicialmente, aplicou-se o filtro categórico na variável UF, selecionando apenas os estados da região Centro-Oeste: Mato Grosso, Mato Grosso do Sul, Distrito Federal e Goiás. Para o agrupamento dos dados, foram escolhidas as variáveis Ano e Trimestre. Diferentemente do primeiro exemplo, optou-se por não incluir a contagem de pessoas nos totais, e foi configurada a agregação para calcular a média da variável V4039 (horas trabalhadas). O arquivo CSV gerado apresentou as colunas Ano, Trimestre e a média das horas trabalhadas (V4039\_media). A figura abaixo apresenta uma visualização gerada a partir desses dados:



**Figura 8** - Horas trabalhadas por semana no Centro-Oeste

## 5 CONCLUSÃO

Este trabalho apresentou o desenvolvimento da biblioteca pnadcontinua, uma ferramenta destinada a simplificar o acesso aos microdados da PNAD Contínua, um recurso valioso para estudos econômicos e sociais no Brasil. Por meio de sua interface gráfica intuitiva, funcionalidades práticas e tecnologias eficientes, a ferramenta possibilita que usuários

realizem consultas, agregações e análises de forma eficiente, superando desafios relacionados ao grande volume de dados e à complexidade do formato original dos arquivos.

A disponibilização da biblioteca no GitHub e no PyPI amplia seu alcance e incentiva sua adoção pela comunidade acadêmica e outros profissionais interessados na exploração desses dados. Ao automatizar tarefas que antes demandavam maior esforço técnico, a pnadcontinua se consolida como uma contribuição significativa para o aprimoramento de análises baseadas nos micrdados da PNAD Contínua, promovendo acessibilidade e eficiência no uso dessa base de dados.

## 6 REFERÊNCIAS

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **PNAD Contínua: O que é?** Disponível em:

<https://www.ibge.gov.br/estatisticas/sociais/trabalho/17270-pnad-continua.html?=&t=o-que-e>.

Acesso em: 10 nov. 2024.

OSÓRIO, Rafael Gomes. **Sobre a montagem de um modelo de identificação longitudinal na PNAD Contínua.** Brasília: Ipea, 2023. Disponível em:

<[https://repositorio.ipea.gov.br/bitstream/11058/11177/1/bmt\\_73\\_sobre\\_montagem.pdf](https://repositorio.ipea.gov.br/bitstream/11058/11177/1/bmt_73_sobre_montagem.pdf)>.

Acesso em: 10 nov. 2024.

POLARS. Polars: **Fast DataFrames in Rust and Python.** Polars. Disponível em:

<https://pola.rs/#:~:text=Polars%20is%20an%20open%2Dsource,expressive%20and%20easy%20to%20use>. Acesso em: 10 nov. 2024.

THE APACHE SOFTWARE FOUNDATION. *Apache Parquet: Columnar storage for analytics.* Apache Software Foundation. Disponível em: <https://parquet.apache.org/>. Acesso em: 10 nov. 2024.