

Trabalho Final

Coleta, Preparação e Análise de Dados

Prof. Luan Garcia



Introdução

Este trabalho extensionista consiste em coleta, preparação e análise de dados governamentais abertos disponíveis através do DATASUS.

Esta atividade deve ser realizada em grupos de 2 até 4 pessoas. Não serão permitidas entregas individuais.

Entrega e Apresentação

Por ser uma atividade extensionista, toda produção do trabalho deve ser disponibilizada de forma aberta. Para isto, cada grupo deve ter um repositório no Github contendo seu código e/ou notebooks utilizados para o pré-processamento e análise de dados.

O que deve ser entregue:

1. Link para Github aberto contendo código e notebooks utilizados no trabalho.
2. Dashboards.
3. Relatório textual descrevendo as bases de dados escolhidas, técnicas de integração, limpeza e transformação utilizadas, conclusões obtidas a partir das análises realizadas, além de referências utilizadas.

Local de entrega: Moodle.

Data de entrega: 26/11/2023

Data de apresentação: 27/11/2023 e 29/11/2023. A ordem de apresentação ocorrerá por sorteio.

Descrição do Trabalho

Você deve escolher uma ou mais bases de dados abertos do DATASUS, pré processá-las, e criar uma visualização para esses dados em formato de *dashboard* que dê suporte à análise de alguma questão de pesquisa a ser respondida com os dados. O portal do DATASUS pode ser acessado no link a seguir:

<https://datasus.saude.gov.br/>

Todas as bases de dados disponíveis podem ser encontradas no portal, e seu download pode ser realizado utilizando o sistema TABNET:

<https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>

Há também algumas bibliotecas disponíveis na internet que criam uma interface para download e pré-processamento dos dados, como o PySUS, uma biblioteca em Python para lidar com dados do

Trabalho Final

Coleta, Preparação e Análise de Dados



Prof. Luan Garcia

DATASUS. Neste caso, não são todas as bases que estão disponíveis para acesso. Abaixo segue uma lista dos sistemas os quais podem ser acessados pelo PySUS:

- **SINAN**: Sistema de Informação de Agravos de Notificação.
- **SINASC**: Sistema de Informação sobre Nascidos Vivos.
- **SIM**: Sistema de Informação sobre Mortalidade.
- **SIH**: Sistema de Informações Hospitalares
- **SIA**: Sistema de Informações Ambulatoriais

Mais informações sobre o PySUS podem ser encontradas no seu repositório:

<https://github.com/AlertaDengue/PySUS>

O uso do PySUS não é obrigatório, todo trabalho pode ser realizado realizando o download das bases diretamente do portal DATASUS.

Ferramentas Sugeridas

Sugere-se o uso das ferramentas a seguir:

- Bibliotecas pandas, scikit-learn e seaborn para o processamento das bases em Python.
- PySUS para acesso a algumas das bases diretamente em Python.
- Tableau ou PowerBI para os dashboards.

Sugestões de Trabalhos

Seu grupo é livre para escolher questões de pesquisa e bases de interesse no DATASUS. Entretanto, segue abaixo algumas sugestões de trabalhos.

- Com a base SINAN, investigar quais regiões do Brasil possuem mais casos de picadas por animais peçonhentos.
- Com a base SINAN, investigar a evolução temporal de casos de dengue nas diferentes regiões do país.
- Com a base SINASC, investigar se existe alguma relação entre partos por cesariana e peso das crianças.
- Com a base SIM e base E-SUS, investigar se existe alguma modificação no número de mortes antes, durante e depois da pandemia, e se existe alguma relação com a vacinação para COVID.

Recursos Interessantes

Abaixo, alguns recursos complementares interessantes.

Trabalho Final

Coleta, Preparação e Análise de Dados



Prof. Luan Garcia

- Documentação PySUS:
<https://pysus.readthedocs.io/en/latest/>
- Artigo no médium sobre download e pré-processamento de dados sobre violência do SINAN:
<https://medium.com/psicodata/download-e-pr%C3%A9-processamento-de-dados-do-sinan-datasus-sobre-viol%C3%Aancia-com-python-ddf46ff9a3f5>
- Artigo no médium sobre investigações de suicídios no Brasil utilizando a base SIM:
<https://medium.com/psicodata/baixando-e-processando-dados-do-datasus-sobre-suic%C3%ADdio-com-python-656afa17f6ad>
- Repositório com diversas análises em notebooks sobre COVID-19:
https://github.com/ceciliasilvads/analises_covid19
- Microdatasus: pacote em R para interface com bases do DATASUS.
<https://github.com/rfsaldanha/microdatasus>