

Relatório: trabalho Final Coleta de Dados

Nomes: Gabriel Mattos, Lorenzo Battistela, Rodrigo Roth e Vinícius Pedroso

1. Introdução

Este relatório descreve a criação e a funcionalidade de um dashboard desenvolvido para monitorar os dados da dengue nas cidades do Rio Grande do Sul. O objetivo do dashboard é fornecer uma análise sobre a situação da dengue na região metropolitana do Rio Grande Do Sul, facilitando a visualização do problema e os locais e condições que podem influenciar a proliferação da doença.

2. Objetivo do Dashboard

O principal objetivo do dashboard é:

- Monitorar a quantidade de casos de dengue nas cidades do Rio Grande do Sul.
- Identificar tendências temporais e geográficas da doença.
- Facilitar a visualização dos dados.

3. Fontes de Dados

Os dados utilizados no dashboard foram obtidos através dos sites “DataSus” disponibilizado pelo professor no enunciado do trabalho e também disponível no link abaixo:

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

4. Estrutura e Funcionalidades do Dashboard

O dashboard foi criado a partir de um “[template](#)” disponibilizado no site “powerBI” dividido em duas seções. As funcionalidades incluem:

4.1. Visão Geral dos Casos

- **Gráfico de Barras:** Exibe a quantidade de casos por cidades
- **Função de Limitar Datas:** Limita as datas para obtermos uma noção de diferentes períodos de tempo
- **Mapa de calor:** Mostra a distribuição geográfica dos casos, permitindo identificar as áreas afetadas.

4.2. Algumas informações referentes aos estados

- **Temperatura**
- **Número Populacional**

- **Número de Casos**

4.3. Estatísticas de cada cidade constatada no DataSet

- **Média anual de temperatura**
- **Média anual de umidade**
- **Temperatura por cidade**
- **Casos por ano**

5. Código do trabalho

Para efetuar a extração, pré-processamento e análise dos dados, o grupo utilizou as seguintes bibliotecas de Python: “BeautifulSoup”, “numpy”, “pandas”, “matplotlib”, “seaborn” e “sklearn”. Utilizamos “webscrapping” para filtrar os nomes das cidades da região metropolitana de Porto Alegre, como consta na função “get_metropolitan_cities”. Após a aquisição dos nomes das cidades, usamos a função “get_city_data” para cada cidade da lista, nela baixamos o “Dataset” sobre a dengue, retiramos as colunas indesejadas, concatenamos os dados das semanas em dados de meses e corrigimos alguns dados faltantes na coluna de temperatura e umidade.

Além do comentado, a equipe tentou construir um modelo de regressão linear utilizando a biblioteca “sklearn”, como se encontra na função “create_temporal_series”. Para a criar a regressão, pensamos em tentar correlacionar fatores coadjuvantes com o resultado, isto é, utilizar as colunas que possuem informações sobre temperatura, humidade e população para tentar chegar numa correlação com a quantidades de casos presente nas cidades. Entretanto, o modelo construído e treinado com os dados selecionado no “dataset”, feito para o “DashBoard”, não foi suficiente para construir uma regressão linear que obtivesse acurácia interessante para predizer a quantidade de casos. Como constam os resultados (1), apenas 3,79% da variância dos casos de dengue puderam ser explicados pelas variáveis selecionadas (como consta o parâmetro R-squared). Além disso os coeficientes relatam números dúbios em relação com o que consta a literatura sobre os estudos de proliferação de contaminação, melhor dizendo, os coeficientes não aparentam representar o mesmo grau de importância em relação ao que conhecemos sobre a disseminação do vírus (2), considerando pouco peso para a temperatura e umidade e elevando a correlação para a quantidade de pessoas (não necessariamente contaminadas).

6. Conclusão

O dashboard de monitoramento da dengue é uma ferramenta que pode ser utilizada para a gestão da saúde pública no Rio Grande do Sul. Ele pode oferecer uma visão detalhada e atualizada da situação da doença, permitindo com que alguma tomada de decisão seja conduzida com mais efetividade. Com a

continuidade do monitoramento e a atualização dos dados, a incidência de dengue pode ser controlada e reduzida, assim ajudando a proteger a saúde da população.

7. Referências

<https://www.hashtagtreinamentos.com/templates-gratuitos-do-power-bi>

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

<https://powerbi.tips/product/layouts-splash-o-red/>

link do dashboard:

<https://app.powerbi.com/view?r=eyJrljoiMjc5ZTg3ZWYtMzQ3Yy00YjUwLWEwYTctZTJkNTZlYmZkOTJmliwidCI6ImY4OTRjOWJjLTNmODUtNGU0Yi05MDgyLTU1NmVmYTNmMTc0YyJ9>

github do trabalho : <https://github.com/Lorenzobattistela/coleta-trabalho-2>

(1)(output do modelo de regressão linear):

“

Mean Squared Error: 4369.370393381743

R-squared: 0.03785158393836219

Coefficient

temperatura 0.120945

umidade 0.571040

pop 5.316319

”

(2):

[https://www.scielo.br/j/cenf/a/5RNNYq6ZPwzpfGbFwSgK9nd/#:~:text=Houve%20correlação%20positiva%20significativa%20\(](https://www.scielo.br/j/cenf/a/5RNNYq6ZPwzpfGbFwSgK9nd/#:~:text=Houve%20correlação%20positiva%20significativa%20()