



FUNDAÇÃO EDSON QUEIROZ
UNIVERSIDADE DE FORTALEZA - UNIFOR
CENTRO DE TECNOLOGIA E ENGENHARIA
CURSO MBA EM ENGENHARIA DE DADOS

TRABALHO FINAL PARA DISCIPLINA: WEB MINING & CRAWLER SCRAPPING

FELIPE ALVES DA SILVA

Fortaleza-CE

2024

FELIPE ALVES DA SILVA

TRABALHO FINAL PARA DISCIPLINA: WEB MINING & CRAWLER SCRAPING

Trabalho final apresentado ao curso de MBA em engenharia de dados da Universidade de Fortaleza, para obtenção de nota final para a disciplina de web mining & crawler scrapping lecionado pelo professor Alex Lima.

Fortaleza-CE

2024

SUMÁRIO

SUMÁRIO3

1 DETALHAMENTO DO PROJETO4

2 REFERÊNCIAS6

LINK PARA NOTEBOOK NO GITHUB:

[HTTPS://GITHUB.COM/FELIPEALVSS/PROJETO_WEB_SCRAPPING/BLOB/MAIN/PROJETO%20FINAL%20WEB%20MINING.IPYNB](https://github.com/felipealvss/projeto_web_scrapping/blob/main/projeto%20final%20WEB%20MINING.ipynb)

1 DETALHAMENTO DO PROJETO

A fonte de dados escolhida para realização do projeto foi o site IntegraSUS (link na parte REFERÊNCIAS), onde escolhi obter dados do cenário das arboviroses no Ceará, mais especificamente os casos de Dengue. Para extração dos dados, foram escolhidas as ferramentas *Selenium* e o *Pandas* para conexão no site, extração e ajuste dos dados.

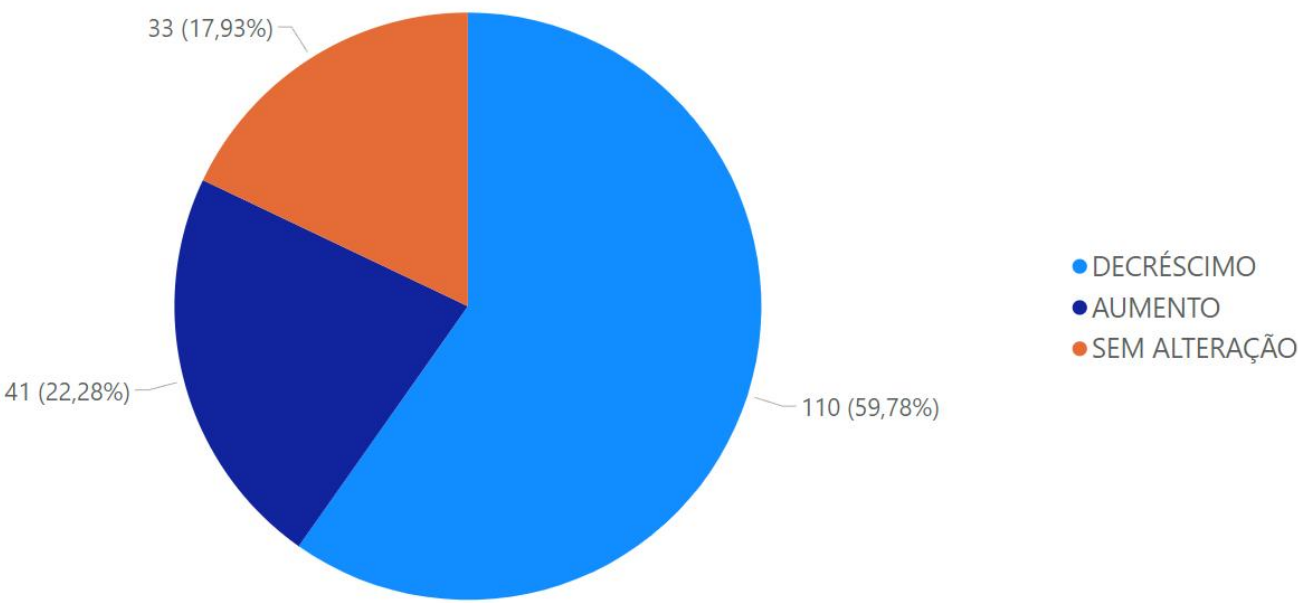
Sobre a execução do Jupyter Notebook: A estrutura do trabalho em si foi deixada de forma simplificada e objetiva, de forma que para executar os processos de extração, normalização, exportação e análise dos dados basta realizar a execução do Jupyter Notebook célula após célula para que o resultado final possa ser gerado (ou se preferir, selecionar a opção de “executar tudo”).

Em relação ao tema escolhido, a pergunta que originou o objetivo da análise escolhida foi: Se vê atualmente uma taxa alta de chuvas ocorrendo por todo o Ceará o que está diretamente ligado a ocorrência de surgimento de focos para o mosquito *Aedes Aegypti*. Tendo isso em mente, a análise busca fazer uma relação entre os dados medidos do ano de 2023 com os dados atuais do ano de 2024 e conferir se em relação apenas nesses 4 meses iniciais houve um aumento percentual no casos de dengue nos municípios do Ceará.

Detalhando o processo realizado, foi feita a extração dos dados do site de acordo por ano, e após a geração dos dados, os mesmos foram extraídos para arquivos .csv; em seguida foi feita uma comparação entre os anos e gerados valores percentuais para cada município e armazenados o dados também em um arquivo .csv; E por último, foi estruturado uma relação entre os dados para gerar um dado unificado e ter uma visualização fácil dos dados para cada município. As informações descritas tanto são possíveis ser visualizadas nos arquivos exportados quanto no DataFrame que é gerado nas etapas finais do Jupyter Notebook.

Para resumir o resultado da análise de forma objetiva, informo em seguida uma imagem de um gráfico aonde se chega a seguinte conclusão: Através de conversas informais, o assunto principal é que está a ocorrer uma grande quantidade de casos de dengue, porém analisando os dados o resultado obtido não é absoluto: Se comparado ao ano de 2023, aproximadamente 60% dos municípios apresentaram um decréscimo de casos confirmados e apenas cerca de 22% dos municípios apresentaram um aumento no número de casos (a 1a imagem a seguir apresenta o agrupamento de casos confirmados por município, e a segunda apresenta as cidades com maiores quantidades de casos no ano de 2024).

CENÁRIO GERAL DOS MUNICÍPIOS DE TODO CEARÁ:



MUNICÍPIOS COM MAIORES VALORES DE 2024 (YTD):

Município	Soma de Confirmados 2023	Soma de Confirmados 2024	Soma de Confirmados %
BREJO SANTO	1781	891	-49,97
FORTALEZA	4981	668	-86,59
TIANGUA	808	158	-80,45
PORTEIRAS	234	127	-45,73
JUAZEIRO DO NORTE	65	98	50,77
CAUCAIA	456	81	-82,24
CRATO	154	44	-71,43
VICOSA DO CEARA	46	38	-17,39
BARROQUINHA	4	28	600,00
TAUA	40	28	-30,00
GRACA	5	27	440,00
JIJOCA DE JERICOACOARA	83	23	-72,29
PACAJUS	31	23	-25,81
ARACATI	71	21	-70,42
ABAIARA	7	20	185,71
LIMOEIRO DO NORTE	492	20	-95,93
MARACANAU	88	20	-77,27
SOBRAL	161	20	-87,58

LINK PARA NOTEBOOK NO GITHUB:

[HTTPS://GITHUB.COM/FELIPEALVSS/PROJETO_DADOS_IBGE/BLOB/MAIN/DESAFIO%20IBGE.IPYNB](https://github.com/felipealvss/projeto_dados_ibge/blob/main/desafio%20ibge.ipynb)

2 REFERÊNCIAS

Site selecionado para extração de dados:

<https://integrasus.saude.ce.gov.br/#/indicadores/indicadores-vigilancia-saude/cenario-arbovirose>

Links selecionados para apoio na construção:

<https://awari.com.br/porcentagem-python-aprenda-a-calcular-porcentagens-usando-python/>

<https://pt.stackoverflow.com/questions/434045/como-converter-valores-em-porcentagem-de-uma-coluna-total-em-pandas>