

# Teste de Conhecimento em Web Scraping e Análise de Dados com Python

Olá! Parabéns por ter conseguido chegar até aqui. Agora gostaríamos de testar os seus conhecimentos e sua capacidade de pesquisar e desenvolver soluções.

Este teste tem como objetivo avaliar suas habilidades e conhecimentos relevantes para a vaga em questão. Responda às questões da melhor forma possível e não se preocupe se não conseguir responder completamente ou todas as questões.

## Questão 1

Utilizando as bibliotecas Requests e BeautifulSoup, escreva um código em Python que faça o web scraping do título, up votes e do link das três primeiras postagens do subreddit r/programming no Reddit.

O algoritmo deve guardar essas informações de forma tabular em uma planilha excel ou csv.

## Questão 2

Considere que você tenha feito o web scraping de uma API que contém o resultado de uma pesquisa em um e-commerce. A resposta da API está em formato json no arquivo api\_response.json enviado junto ao desafio.

O arquivo possui inúmeras informações utilizadas pelo site para mostrar os resultados, mas seu trabalho é extrair uma lista com as ofertas resultantes da pesquisa. O objetivo é extrair os atributos principais das ofertas.

Exemplo de pesquisa: [https://www.climario.com.br/geladeira?\\_q=geladeira&map=ft](https://www.climario.com.br/geladeira?_q=geladeira&map=ft)

List of:

- offer\_link -> link da oferta
- image\_link -> link da imagem do produto
- price -> preço da oferta
- title -> título da oferta

## Questão 3:

Usando Python e quaisquer outras bibliotecas que julgar necessárias. Faça a captura das informações de um produto do site <https://www.magazineluiza.com.br/> e estruture as informações num json.

Exemplo de oferta:

<https://www.magazineluiza.com.br/aparador-pelos-philips-one-blade-qp1424-10-philips-novo/p/237634500/pf/papp/>

Os atributos que deverão ser capturados são:

- title
- stock\_availability
- price

Dica: além do request diretamente na página do produto, você também pode explorar as requisições que o site faz, a fim de descobrir alguma API, ou então utilizar alguma biblioteca para renderizar o conteúdo da página, como Selenium.

## Questão 4:

Nesta tarefa, você recebeu a responsabilidade de conduzir uma análise detalhada dos dados de compras de um website. Os dados estão fornecidos no formato JSON (dados\_compra.json).

O website em questão permite que os usuários façam login em suas contas pessoais e realizem compras enquanto exploram a lista de produtos disponíveis. Cada produto possui um valor de venda associado. Além disso, informações demográficas, como idade e sexo, foram coletadas para cada usuário e estão presentes no arquivo JSON. Seu objetivo é fornecer uma análise abrangente do comportamento de compra dos consumidores.

Nesta missão, você irá conduzir a análise de compra dos consumidores utilizando a biblioteca Pandas da linguagem Python. Seu relatório final deve contemplar os seguintes itens:

## Exploração dos Dados

- Carregue os dados JSON utilizando a biblioteca Pandas.
- Examine as primeiras linhas do conjunto de dados para entender sua estrutura.
- Verifique a presença de quaisquer valores ausentes nos dados.
- Identifique a quantidade total de compras realizadas.

### Análise de Compras

- Calcule a média, o valor mínimo e máximo gasto por compra.
- Determine o produto mais caro e o produto mais barato.

### Segmentação por Gênero

- Analise a distribuição de gênero entre os consumidores.
- Calcule o valor total gasto em compras por gênero.

### Visualização de Dados

- Utilize gráficos adequados (como histogramas, gráficos de barras, etc.) para ilustrar os insights obtidos. Fique à vontade.
- Certifique-se de adicionar títulos e rótulos aos gráficos para facilitar a compreensão.

Como considerações finais:

- Seu script deve funcionar para o conjunto de dados fornecido.
- Fique à vontade para incrementar as análises.

## Entrega

Ao finalizar o desafio publique o código no Github da forma que achar mais conveniente e nos envie o link. Lembre de deixá-lo público para que possamos ver.

Você tem até a quarta-feira 04/11 às 23h59 para enviar o link com o repositório do Github no email: [paco.mora@hooklab.com.br](mailto:paco.mora@hooklab.com.br)

Bom desafio!