

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.

Questão 1

Utilizando a biblioteca Requests 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 informações sobre filmes em um site de avaliações de filmes. Você obteve os seguintes dados para três filmes:

```
filmes = [
    {"titulo": "O Senhor dos Anéis", "ano": 2001, "avaliacao": 8.8},
    {"titulo": "Matrix", "ano": 1999, "avaliacao": 9.3},
    {"titulo": "Interestelar", "ano": 2014, "avaliacao": 8.6}
]
```

Calcule e apresente:

- A média das avaliações dos filmes.
- O título do filme com a maior avaliação.
- O ano de lançamento do filme com a menor avaliação.

Questão 3

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 16/08 às 23h59 para enviar o link com o repositório do Github no email: paco.mora@hooklab.com.br

Bom desafio!