**Teste realizado sobre a ferramenta Scrapy**

Objetivo:

Realizar a raspagem dos dados de um produto específico da página da mericanas.com, e chegar a exportação para um arquivo JSON com os dados extraídos.

Configuração do projeto:

O projeto do teste foi desenvolvido com Python 3.8.2 e framework Scrapy 1.8.0.

Uma coisa ser comiserada e que a configuração de DEFAULT\_REQUEST\_HEADERS do projeto foi alterado para atender a necessidade do meu ambiente uma vez que a pagina alvo não estava aceitando o agente informado por default no framework, então e possível que em outros ambientes estas configurações possam ter que ser mudadas.

Outra configuração feita foi da variável de ambiente FEED\_EXPORT\_ENCODING, para UTF-8, para que ao exportar para o arquivo JSON os caracteres acentuados, ficassem legíveis, uma vez que o comportamento default da ferramenta e faz a exportação com números seguros por rações históricos conforme a documentação, então como este requisito estava aberto optei por usar o UTF-8.

Decisões de projeto:

O avaliador também poderá notar que optei por fazer o tratamento dos dados usando ItemLloader e não Item Pipiline, uma vez que a documentação do framework sugere que esta e uma melhor pratica para fazer o tratamento individual dos campos. Sendo o Pipeline, mas adequado para tratamentos mais robustos como duplicidade de dados, validação de estruturas, descarte, envio dos dados para uma base de dados ou até mesmo a orquestração de Pipelines que executem estas diversas ações em um flow. Porem para fins de apresentação está desabilitado no projeto um Pipiline que deve fazer o download da imagem do projeto, precisando para isso ter a permissão para realizar esta operação e também de criar a pasta onde o arquivo esta salva. Para habilita-lo basta remover os comentários em ITEM\_PIPELINESr

Execução:

Para executar o projeto só executar o comando abaixo

scrapy crawl produtospider -o produto.json

Referências:

<https://docs.scrapy.org/en/latest/topics/loaders.html>

<https://docs.scrapy.org/en/latest/topics/item-pipeline.html>

<https://docs.scrapy.org/en/latest/topics/feed-exports.html#feed-export-encoding>