Answers:

# Python Exercises

## Test #1

Python and CSV files attached separately

## Test #2

### 2.1

Import re

product\_count\_text = ‘381 Products found’

ans) product\_count\_int = int(re.findall('(\d.\*\d)\sProducts', product\_count\_text)[0])

### 2.2

generic\_urls = []

for url in generic\_urls:

ans) special\_sequence = re.findall('.com.\*/(.\*)-',url)

### 2.3

import json

import re

product\_storage\_string\_raw = ‘given’

ans) product\_storage\_parsed = json.dumps({"price":re.findall('price":([0-9].\*?),"name',product\_storage\_string\_raw),"name":re.findall('name":"(.\*?)","id', product\_storage\_string\_raw),"id":str(re.findall('id":"(.\*?)","position',product\_storage\_string\_raw)),"position":re.findall('position":(.\*?),"list',product\_storage\_string\_raw),"list":re.findall('list":"(.\*?)","category',product\_storage\_string\_raw),"category":re.findall('category":"(.\*?)"',product\_storage\_string\_raw),"brand":re.findall('brand":"(.\*?)"},',product\_storage\_string\_raw)})

product\_storage\_json = json.loads(product\_storage\_parsed)

# Technical test – Xpaths

## Test #1

Scraping single information from a page onto a column

### Obtained output:

Graphical user interface, application, Word

Description automatically generated

Main XPath: //div[3]/div[5]/div/div[2]/div/a

Category: .

## Test #2

Scraping multiple information from a page onto multiple columns

### Obtained output:

Table

Description automatically generated

Main XPath: //\*[@id="product-listing-container"]/form[2]/ul[1]/ul/li

Product Name: .//article//div//h4//a

Was Price: .//article//div//div//div//div//span[1]

Now Price: .//article//div//div//div//div//span[2]

Product URL: .//article//figure//a//@href

Image URL: [.//article//figure//a//img//@src](mailto:.//article//figure//a//img//@src)

## Test #3

Navigating through menu and scraping its information

### Obtained output:

Graphical user interface, text

Description automatically generated

Main XPath: //div[@id='hmenu-content']/ul[position() > 1]//li//a[not(@class='hmenu-item hmenu-compressed-btn' or @class='hmenu-item hmenu-expanded-btn' or @class='hmenu-item hmenu-back-button')]

Category: ../../li[2]

Subcategory: .

Link: @href