Yaisah Granillo   04/10/2019
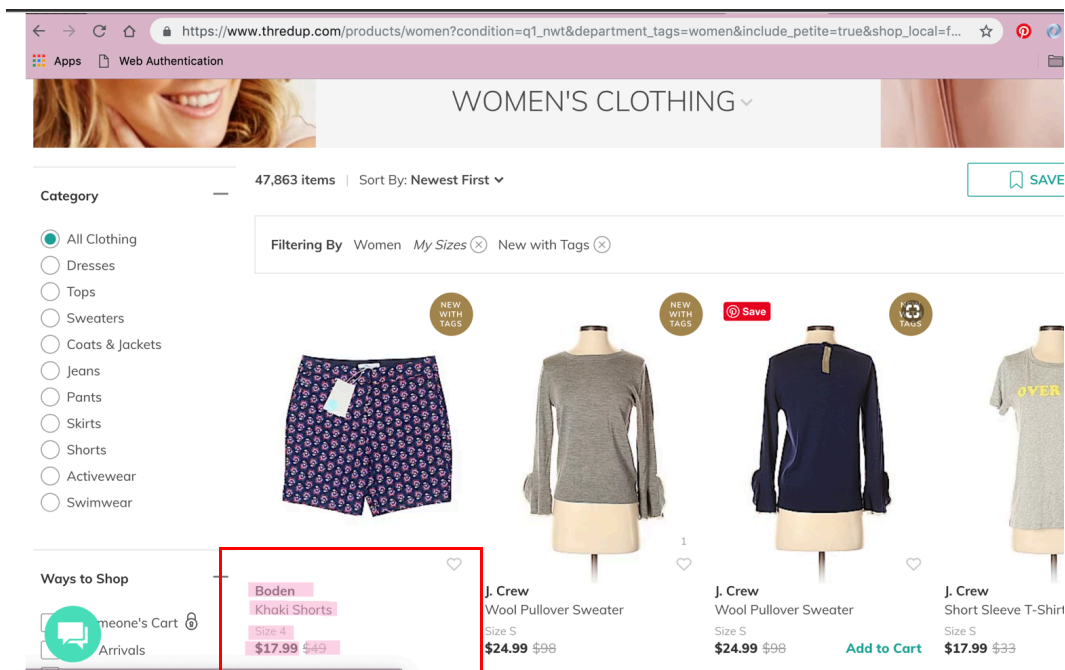Blockchain and Enterprise Data

Assignment #3B – Web Crawling

Part 1: Website Identification and Mapping

ThredUp: With ThredUP, there is no need to shop for high quality brands at retail price nor a need to worry about peer to peer shopping (i.e. Poshmark, eBay, etc.). This is a reliable secondhand shop website for women and children. If you don't like what you purchased, you simply return it.

URL (Customized to my sizes and filtered to only new products with tags):
https://www.thredup.com/products/women?
condition=q1_nwt&department_tags=women&include_petite=true&shop_local=false&sizing_id
=551%2C765%2C778%2C781%2C774%2C791%2C788&sort=Newest+First

Below is the HTML code that I'll be using to scrape only the necessary data for easier shopping!

```
                </div>...  <link  tabindex...  Boden  div>
              </a>
            ▼ <a href="/product/women-boden-blue-khaki-shorts/52359310?
              sizing_id=551%2C765%2C778%2C781%2C774%2C791%2C788">
                <div class="item-title">Khaki Shorts</div>
                <div class="size-name">Size 4</div>
              ▼ <div class="price-line">
                ▼ <div class="formatted-prices">
                  ▶ <span class="formatted-price price-color--regular">…</span>
                  ▶ <span class="formatted-msrp">…</span>
                  </div>
                  <button aria-label="Add to Cart" class="_2qc0J js-modal-trigge
                  atc" type="button">Add to Cart</button>
                </div>
              </a>
              <!-- react-empty: 22044 -->
            </div>
          </div>
        </div>
      ▶ <div class="results-grid-item">…</div>
      ▶ <div class="results-grid-item">…</div>
      ▶ <div class="results-grid-item">…</div>
      ▶ <div class="results-grid-item">…</div>
      ▶ <div class="results-grid-item">…</div>
```

Raw HTML

<div class="results-grid-item"><div class="carousel-results-item item-card listed"><script type="application/ld+json">{"@context":"http://schema.org","@type":"Product","color":"Blue","brand":{"@type":"Brand","name":"Boden"},"description":"Blue. Size 4","sku":"52359310","image":["https://cf-assets-thredup.thredup.com/assets/146386720/retina.jpg","https://cf-assets-thredup.thredup.com/assets/146386755/retina.jpg"],"name":"Khaki Shorts","offers":{"@type":"Offer","availability":"InStock","itemCondition":"UsedCondition","price":17.99,"priceCurrency":"USD","seller":{"@type":"Organization","legalName":"THREDUP"}}}</script><div class="item-card-top"><div><div class="_3F6I2 item-banner item-banner-full-round item-banner-nwt2" style="padding-top: 133.333%;"><div class="_1fkzD"><div class="_1FCYQ" style="opacity: 0; transition: opacity 500ms linear 200ms; background-color: rgba(49, 53, 90, 0.25);"></div><!-- react-text: 22016 --><!-- /react-text --><img alt="Boden Khaki Shorts Size 4" src="https://cf-assets-thredup.thredup.com/assets/146386720/complimentary.jpg" style="opacity: 1; transition: opacity 500ms linear 200ms;"></div></div></div><!-- react-empty: 22017 --><div class="js-modal-trigger-favorite small-favorite-container no-count"> Favorited <!-- react-text: 22021 -->0<!-- /react-text --> times <button aria-label="Add favorite" class="favorite-button f4DZ1"><!-- react-text: 22026 --> <!-- /react-text --><!-- react-text: 22027 -->Favorite Item<!-- /react-text --><!-- react-text: 22028 --> <!-- /react-text --></button></div></div><div class="item-card-bottom"><div role="link" tabindex="0">Boden</div><div class="item-title">Khaki Shorts</div><div class="size-name">Size 4</div><div class="price-line"><div class="formatted-prices"><!-- react-text: 22038 -->$<!-- /react-text --><!-- react-text: 22039 -->17.99<!-- /react-text --><!-- react-text: 22041 -->$<!-- /react-text --><!-- react-text: 22042 -->49<!-- /react-text --></div><button aria-label="Add to Cart" class="_2qc0J js-modal-trigger-atc" type="button">Add to Cart</button></div><!-- react-empty: 22044 --></div></div></div>

Part 2: Database Table Creation

| | Browse | Structure | SQL | Search | Insert | Export | Import | Operations | Triggers |

| | Table structure | Relation view |

#	Name	Type	Collation	Attributes	Null	Default	Extra	Action
1	Brand	varchar(100)	latin1_swedish_ci		No	None		Change Drop Primary Unique Index Spatial Fulltext Distinct values
2	itemtitle	varchar(100)	latin1_swedish_ci		No	None		Change Drop Primary Unique Index Spatial Fulltext Distinct values
3	size	varchar(50)	latin1_swedish_ci		No	None		Change Drop Primary Unique Index Spatial Fulltext Distinct values

phpMyAdmin

← Server: localhost » Database: ygranill » Table: WebcrawlerData3B

Browse Structure SQL Search Insert Export Import Operations Triggers

⚠ Current selection does not contain a unique column. Grid edit, checkbox, Edit, Copy and Delete features are not available. ⓘ

✔ Showing rows 0 - 24 (97 total, Query took 0.0008 seconds.)

SELECT * FROM `WebcrawlerData3B`

Recent Favorites

- information_schema
- ygranill
  - New
  - OwnerPet
  - TEST
  - ThredUp
  - WebcrawlerData3B
    - Columns
      - New
      - Brand
      - itemtitle
      - link
      - MSRP
      - price
      - size
    - YGRANILL_EXAM1
      - Columns
      - Indexes

1 ▾ > >> ☐ Show all Number of rows: 25 ▾ Filter rows: Search this table

+ Options

Brand	itemtitle	size	price	MSRP	link
Lularoe	Cardigan	Size XS	$31.99	$48	https://www.thredup.com/products/women/lularoe?bra...
Boden	Khaki Shorts	Size 4	$17.99	$49	https://www.thredup.com/products/women/boden?brand...
J. Crew	Wool Pullover Sweater	Size S	$24.99	$98	https://www.thredup.com/products/women/j-crew?bran...
J. Crew	Wool Pullover Sweater	Size S	$24.99	$98	https://www.thredup.com/products/women/j-crew?bran...
J. Crew	Short Sleeve T-Shirt	Size S	$17.99	$33	https://www.thredup.com/products/women/j-crew?bran...
Helmut Lang	Casual Dress	Size 2	$135.99	$550	https://www.thredup.com/products/women/helmut-lang...
Mod Bod Swim	Swimsuit Top	Size XS	$2.99	$24	https://www.thredup.com/products/women/mod-bod-swi...
Tevolio	Cocktail Dress	Size 4	$28.99	$89	https://www.thredup.com/products/women/tevolio?bra...
Tommy Bahama	Casual Dress	Size XS	$44.99	$128	https://www.thredup.com/products/women/tommy-baham...
Express	Short Sleeve Blouse	Size S (Petite)	$17.99	$50	https://www.thredup.com/products/women/express?bra...
BCBGeneration	Cocktail Dress	Size 4	$40.99	$128	https://www.thredup.com/products/women/bcbgenerati...
Levi's	Denim Vest	Size S	$32.99	$88	https://www.thredup.com/products/women/levis?brand...
White House Black Market	Shorts	Size 2	$32.99	$68	https://www.thredup.com/products/women/white-house...
Dana Buchman	Long Sleeve Top	Size S	$14.99	$44	https://www.thredup.com/products/women/dana-buchma...
Ann Taylor LOFT	Formal Skirt	Size S	$21.99	$69	https://www.thredup.com/products/women/ann-taylor-...
Broadway & Broome	Casual Skirt	Size 4	$21.99	$69	https://www.thredup.com/products/women/broadway-an...
Jeans By Buffalo	Casual Dress	Size S	$28.99	$65	https://www.thredup.com/products/women/jeans-by-bu...
World Unity	Casual Dress	Size S	$36.99	$60	https://www.thredup.com/products/women/world-unity...
Trina Turk	Cocktail Dress	Size 2	$61.99	$378	https://www.thredup.com/products/women/trina-turk?...
Becca	Jeans	28 Waist	$59.99	$198	https://www.thredup.com/products/women/becca?brand...
White House Black Market	Blazer	Size 4	$36.99	$150	https://www.thredup.com/products/women/white-house...

Part 3: Beautiful Soup Python Program

Running successfuly

C:\Program Files (x86)\Microsoft Visual Studio\Shared\Python36_64\python.exe

```
$24
https://www.thredup.com/products/women/lularoe?brand_name_tags=Lularoe&department_tags=women
Ann Taylor LOFT
Pullover Sweater
Size S
$23.99
$69
https://www.thredup.com/products/women/ann-taylor-loft?brand_name_tags=Ann+Taylor+-+All+Brands&department_tags=women
Lularoe
Short Sleeve T-Shirt
Size XS
$21.99
$24
https://www.thredup.com/products/women/lularoe?brand_name_tags=Lularoe&department_tags=women
GiGi
Active Tank
Size S
$17.99
$78
https://www.thredup.com/products/women/gigi?brand_name_tags=GiGi&department_tags=women
Lularoe
Cardigan
Size XS
$31.99
$48
https://www.thredup.com/products/women/lularoe?brand_name_tags=Lularoe&department_tags=women
Hawke & Co.
Snow Pants
Size S
$117.99
$286
https://www.thredup.com/products/women/hawke-and-co?brand_name_tags=Hawke+%26+Co.+-+All+Brands&department_tags=women
Mink Pink
Casual Pants
Size S
$27.99
$109
https://www.thredup.com/products/women/mink-pink?brand_name_tags=Mink+Pink+-+All+Brands&department_tags=women
Gap Outlet
Long Sleeve Button-Down Shirt
Size S
$28.99
$49
https://www.thredup.com/products/women/gap-outlet?brand_name_tags=Gap+Outlet+-+All+Brands&department_tags=women
```

Complete source code for python project.

```python
#Yaisah Granillo
##
#Don't forget to import the necessary modules
from     import
from         import
```

```python
import
import
import

##Database connection to store the results
                                        "xxx.xxx.xxx"          "xxx"
        "xxx              "xxx"


##website source code – I chose ThredUp filtered to my sizes
      'https://www.thredup.com/products/women?
condition=q1_nwt&department_tags=women&include_petite=true&shop_
local=false&sizing_id=551%2C765%2C778%2C781%2C774%2C791%2C788&so
rt=Newest+First'

                                        'html.parser'


##Get a specific section of a page based on a class, in this
case I want to look at the items in the grid
                                    'results-grid-item'
#print(len(results))

#menuLinks = menuNav2.find_all('a')

for           in
    #Name of the item brand
                              'a'          '_3p5eh'
    print

    #Type of item
                              'div'        'item-title'
    print

    #Size of item
                              'div'        'size-name'
    print

    #current price in the website
                              'span'          'formatted-
price'
    print

    #original price of the product in retail
                              'span'        'formatted-msrp'
    print

    #link to the product
                              'a'          '_3p5eh'  'href'
            "https://www.thredup.com"
    print

##Post the record to desired database
        "INSERT INTO ygranill.WebcrawlerData3B(Brand,
itemtitle, size, price, MSRP, link) VALUES (%s, %s, %s, %s, %s,
%s);"
```

```
##Execute the query


print "Done!"

##close the database connections
```