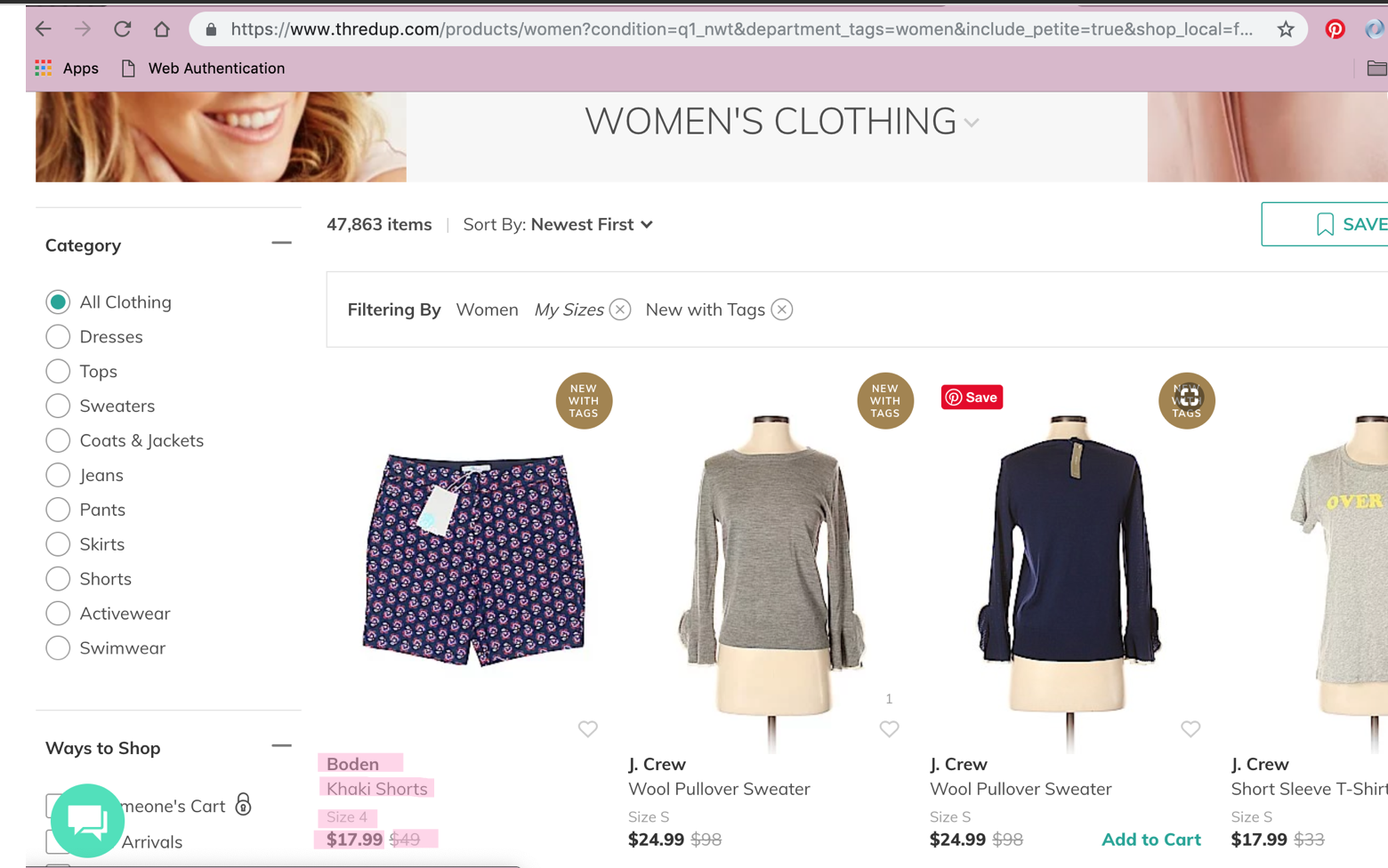
**Assignment #3B – Web Crawling**

# 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>



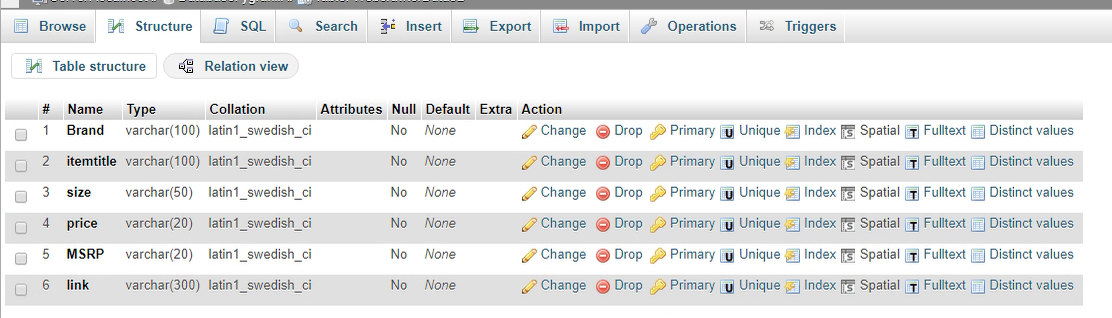
Below is the HTML code that I’ll be using to scrape only the necessary data for easier shopping!

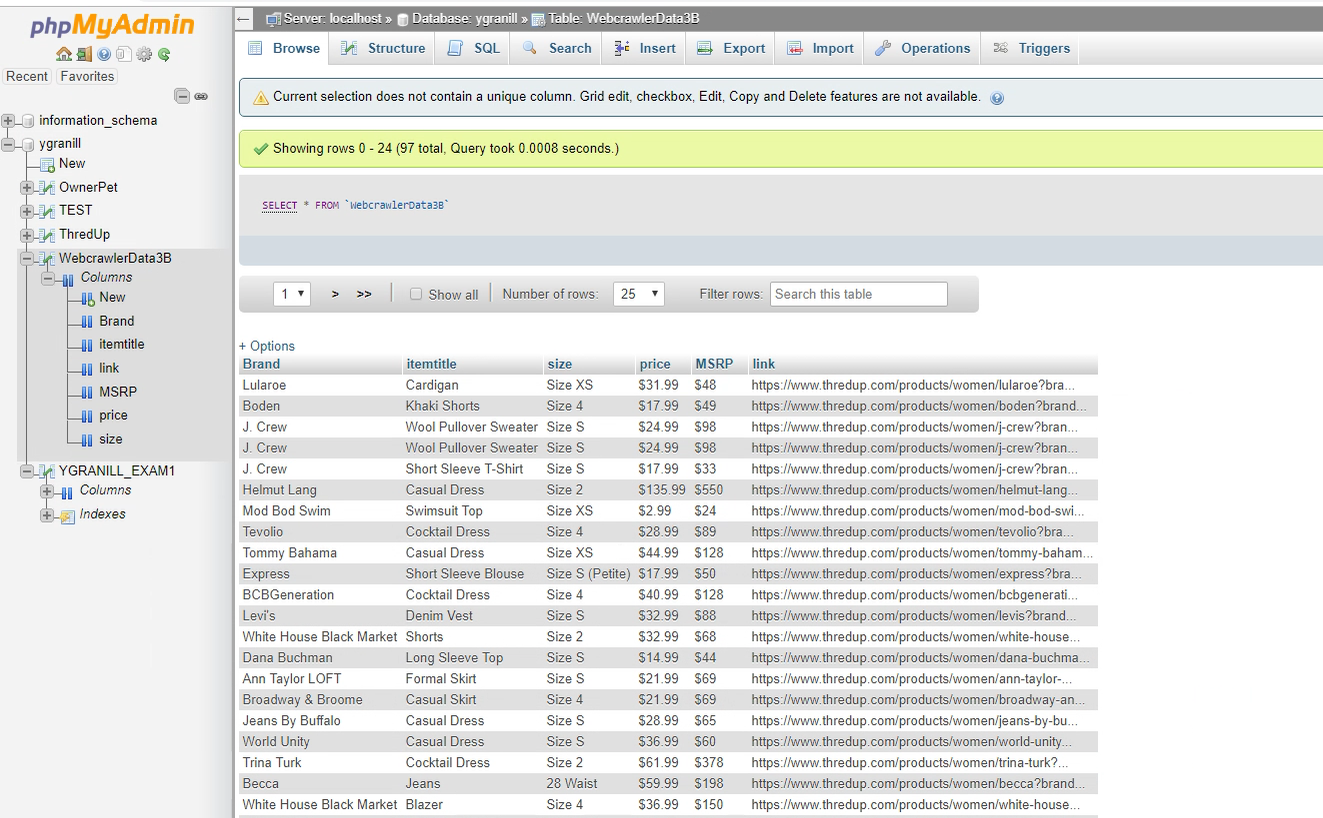


## Raw HTML

<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"><a href="/product/women-boden-blue-khaki-shorts/52359310?sizing\_id=551%2C765%2C778%2C781%2C774%2C791%2C788"><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></a><!-- react-empty: 22017 --><div class="js-modal-trigger-favorite small-favorite-container no-count"><span class="favorite-count hide-favorite-count"><span class="u-visually-hidden"> Favorited </span><!-- react-text: 22021 -->0<!-- /react-text --><span class="u-visually-hidden"> times </span></span><span class="\_1F2hC" aria-label="favorite button"><button aria-label="Add favorite" class="favorite-button f4DZ1"><span class="u-visually-hidden"><!-- react-text: 22026 --> <!-- /react-text --><!-- react-text: 22027 -->Favorite Item<!-- /react-text --><!-- react-text: 22028 --> <!-- /react-text --></span></button></span></div></div><div class="item-card-bottom"><a class="\_3p5eh" href="/products/women/boden?brand\_name\_tags=Boden+-+All+Brands&amp;department\_tags=women&amp;shop\_local=false"><div role="link" tabindex="0">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"><!-- react-text: 22038 -->$<!-- /react-text --><!-- react-text: 22039 -->17.99<!-- /react-text --></span><span class="formatted-msrp"><!-- react-text: 22041 -->$<!-- /react-text --><!-- react-text: 22042 -->49<!-- /react-text --></span></div><button aria-label="Add to Cart" class="\_2qc0J js-modal-trigger-atc" type="button">Add to Cart</button></div></a><!-- react-empty: 22044 --></div></div></div>

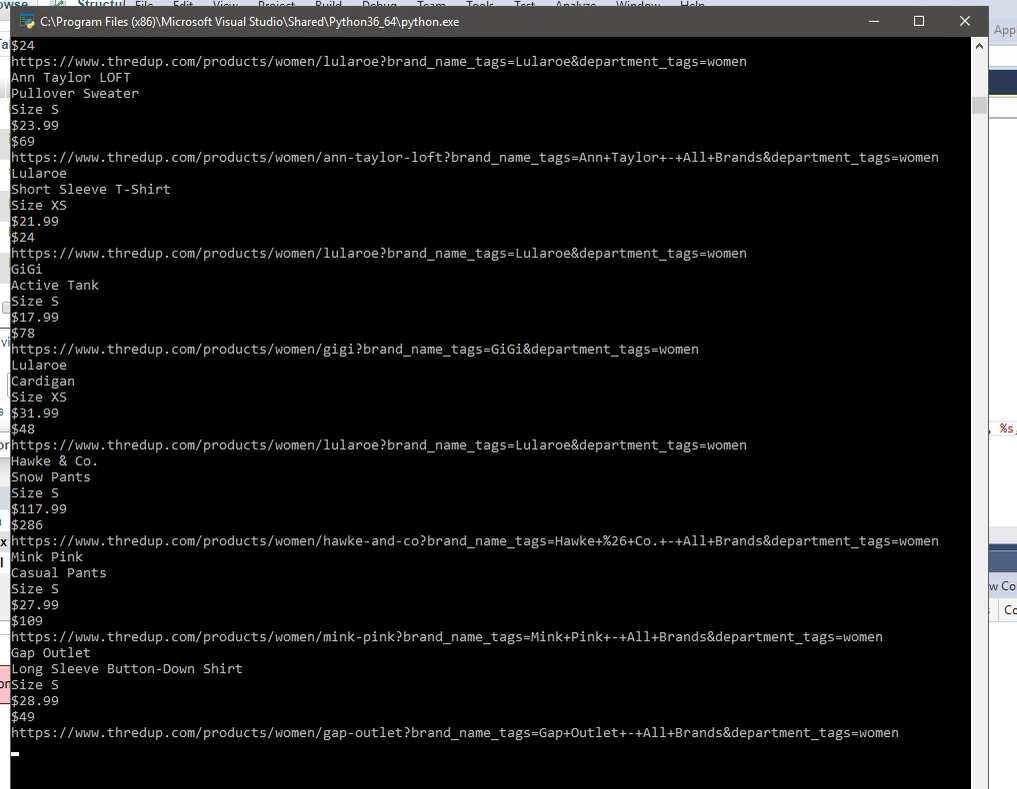
# Part 2: Database Table Creation





# Part 3: Beautiful Soup Python Program

Running successfuly



# Complete source code for python project.

*#Yaisah Granillo*

*##*

*#Don't forget to import the necessary modules*

**from** bs4 **import** BeautifulSoup

**from** requests **import** get

**import** urllib.request, urllib.parse, urllib.error

**import** json

**import** mysql.connector

*##Database connection to store the results*

conn = mysql.connector.connect(host="xxx.xxx.xxx", user="xxx” password="xxx database="xxx")

c = conn.cursor()

*##website source code - I chose ThredUp filtered to my sizes*

url = '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'

response = get(url)

soup = BeautifulSoup(response.text, 'html.parser')

type(soup)

*##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 = soup.find\_all(class\_='results-grid-item')

*#print(len(results))*

*#menuLinks = menuNav2.find\_all('a')*

**for** container **in** results:

*#Name of the item brand*

brand = container.find('a', class\_='\_3p5eh').text

**print**(brand)

*#Type of item*

item = container.find('div', class\_='item-title').text

**print**(item)

*#Size of item*

size = container.find('div', class\_='size-name').text

**print**(size)

*#current price in the website*

price = container.find('span', class\_ = 'formatted-price').text

**print**(price)

*#original price of the product in retail*

msrp = container.find('span', class\_ = 'formatted-msrp').text

**print**(msrp)

*#link to the product*

link = container.find('a', class\_='\_3p5eh')['href']

linkurl = ("https://www.thredup.com"+ link)

**print**(linkurl)

*##Post the record to desired database*

myquery = "INSERT INTO ygranill.WebcrawlerData3B(Brand, itemtitle, size, price, MSRP, link) VALUES (%s, %s, %s, %s, %s, %s);"

arguments = (brand, item, size, price, msrp, linkurl)

*##Execute the query*

c.execute(myquery, arguments)

conn.commit()

**print**("Done!")

*##close the database connections*

c.close()

conn.close()