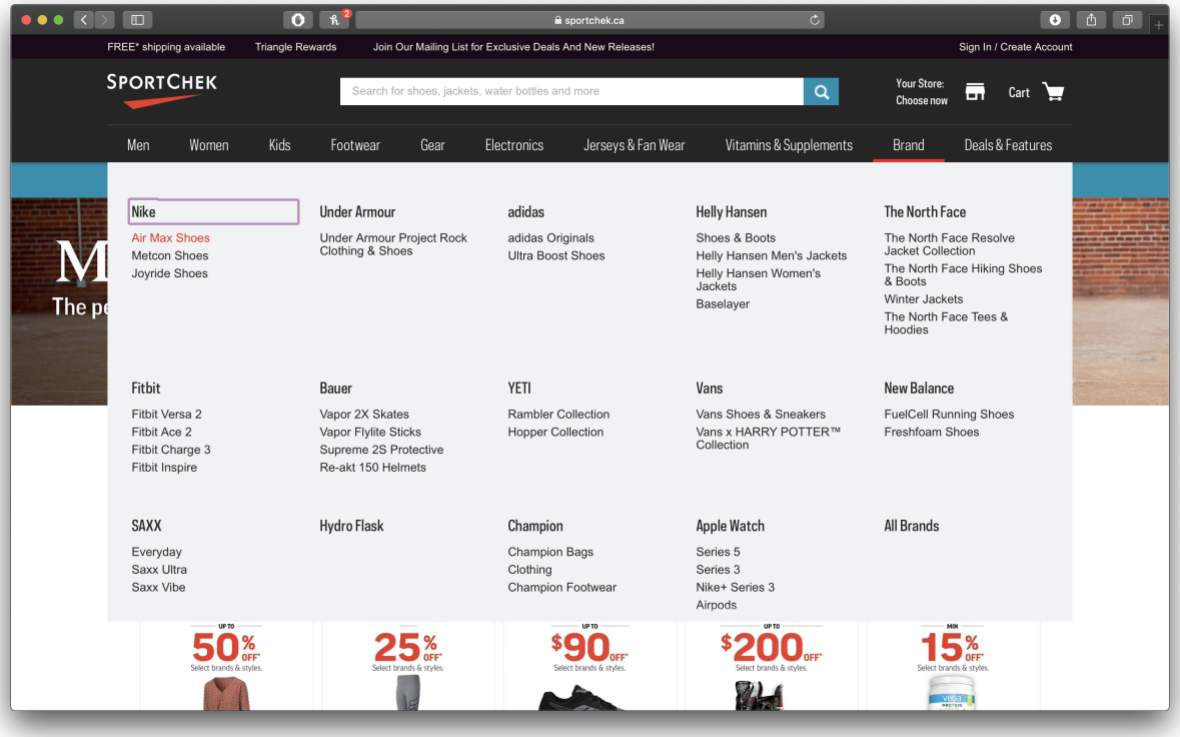
Design of the web scraper

After carefully getting feedback from the professor and reading the assignment I have designed the web scraper to go through three simple levels to retrieve the information which is asked for. The three levels which are being scraped are the home page (landing page), the product list page, and a product page. The reason I am only traversing on these three levels is because the general structure of sportchek contains of the three levels which I previously mentioned.

Levels being scraped

Home page

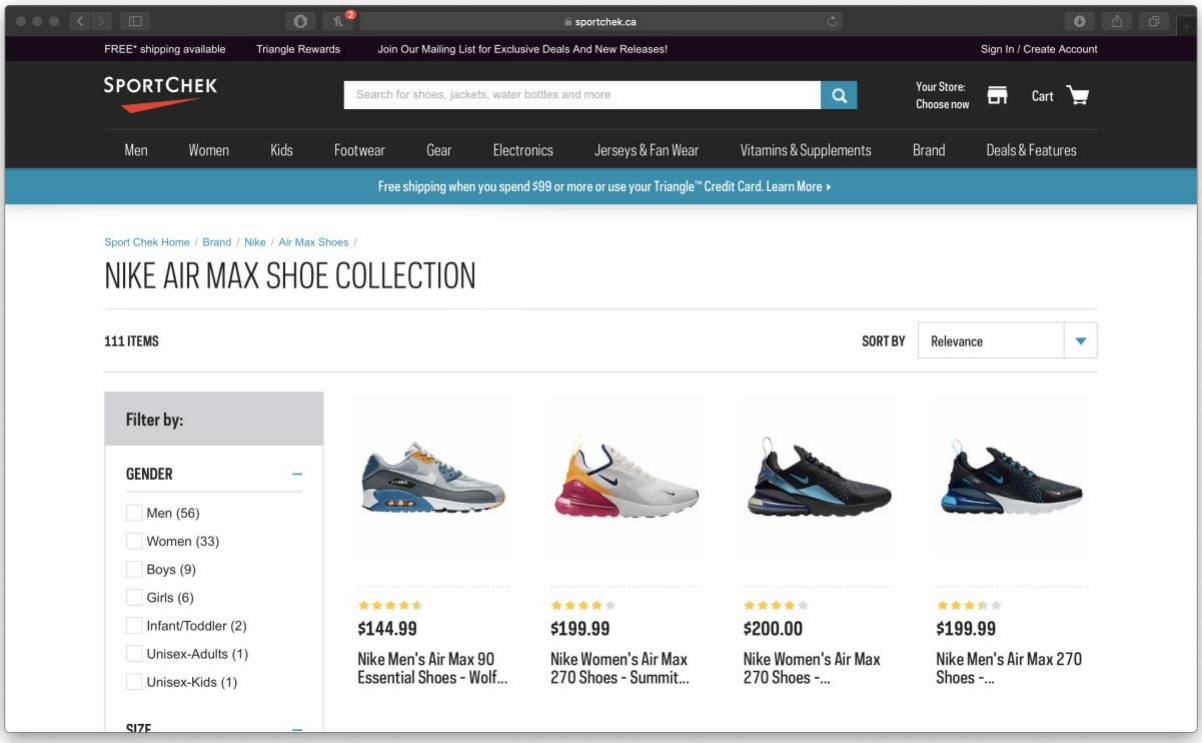
On the home page the code is scraping only the navigation bar as that is where the product list pages can be retrieved through. The first step is to get all the navigation categories name and once we reach the last category we enter its dropdown and get the first link in the first subcategory. In my case it was the categories as shown below. My cursor shows which product list page it goes into.



I could have gone through every single link in the dropdown menu but for the sake of the time taken to run the code I limited the crawler to go to only one product list page. If not the program could go on for an extremely long time.

Product list page

On this page the product items are listed inside a `*ul*` tag and each individual tile is a `*li*` tag. The code takes the href element in the `*a*` tags which are present inside the *li* items and enters them to reach the product page. At the moment I am not going through the pagination and only taking the products from the first page. The screenshot below shows the product list page.

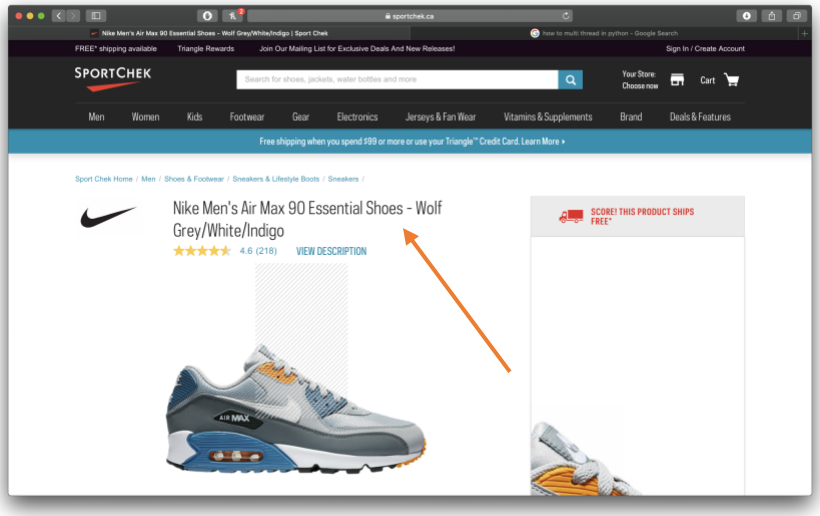


Product page

Once inside the product page the code searches for the breadcrumb at the top of the page. The reason why we take this is so that we can identify what are the categories for that specific product. When outputting the breadcrumb, I skip the first category which is *"Sport Chek Home"* because it does not signify much about what the products. The figure below shows what a breadcrumb is.

Sport Chek Home / Men / Shoes & Footwear / Sneakers & Lifestyle Boots / Sneakers /

The code then takes the product name which is pointed in the figure below.



Lastly, the code retrieves the product number, and the model number of the product from the product description section. This can be seen in the figure below.