**Here are the detailed steps for the data gathering task:**

**Deliverable is the scripts and data in JSON.**

Currently there are 10 e-commerce websites: **Must** **Completely Scrape (every product from each category) the first website (farfetch) in 2 to3 business days.** Based on the priority, you should go with Farfetch first, then BloomingDales and so on.

**First website Have to deliver this week. The URLs are the following:**

**Ahmed:**

1. <https://www.farfetch.com/> (approved & deployed)
2. https://www.newbalance.com.sa/en/
3. <https://tedbaker.sa/> (sample)

Bahram

1. <https://bloomingdales.ae/> (sample)
2. https://sa.hm.com/en/
3. https://theluxurycloset.com/

Fahad

1. <https://ae.diesel.com/> (sample)
2. <https://sandro.sa/> (smple)
3. <https://maje.sa/> (sample)

https://www.bathandbodyworks.com.sa/en/

**And the data fields are:**

**Source ID**: A unique identifier for the source website.

**Product URL**: The URL link to the product page.

**Product Title**: The name of the product.

**Category**: The main and subcategories to which the product belongs.

**Price**: The current price of the product.

**Discount**: Any discount applied to the product, if available.

**Currency**: The currency in which the price is displayed (default: SAR - Saudi Riyal).

**Description**: A detailed description of the product.

**Main Image URL**: URL to our S3 bucket (or similar solution) of the primary product image.

**Other Images URL**: URLs to our S3 bucket (or similar solution) of any additional images for the product.

**Colors**: Available color options for the product.

**Variations**: Different product variations (e.g., sizes, styles, etc.).

**Sizes**: Available size options for the product.

**Other Details**: Any additional product information (e.g., brand, manufacturer).

**Status**: The product's availability status (e.g., “in stock,” “out of stock,” or “discontinued”).

**Number of Items in Stock**: Quantity of items available in stock.

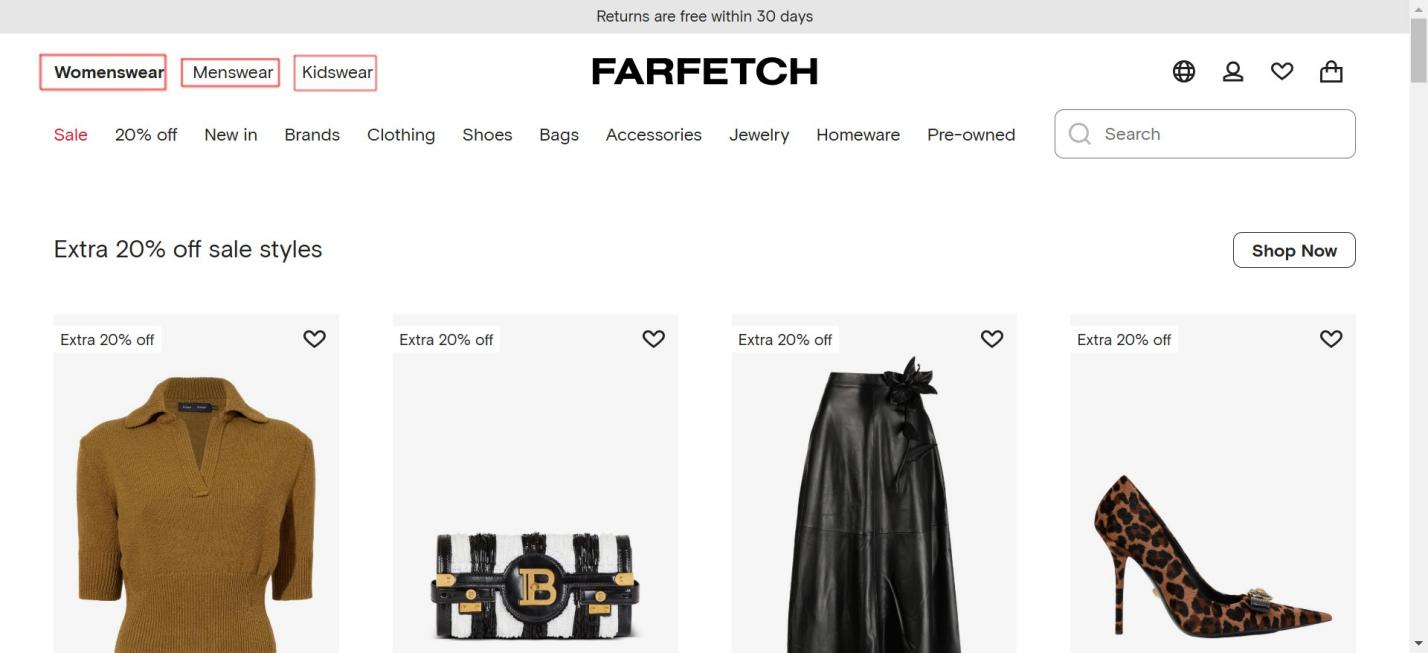
**Last Update**: A timestamp recording the last time the product data was updated.

**Creation Date**: A timestamp indicating when the product data was first captured.

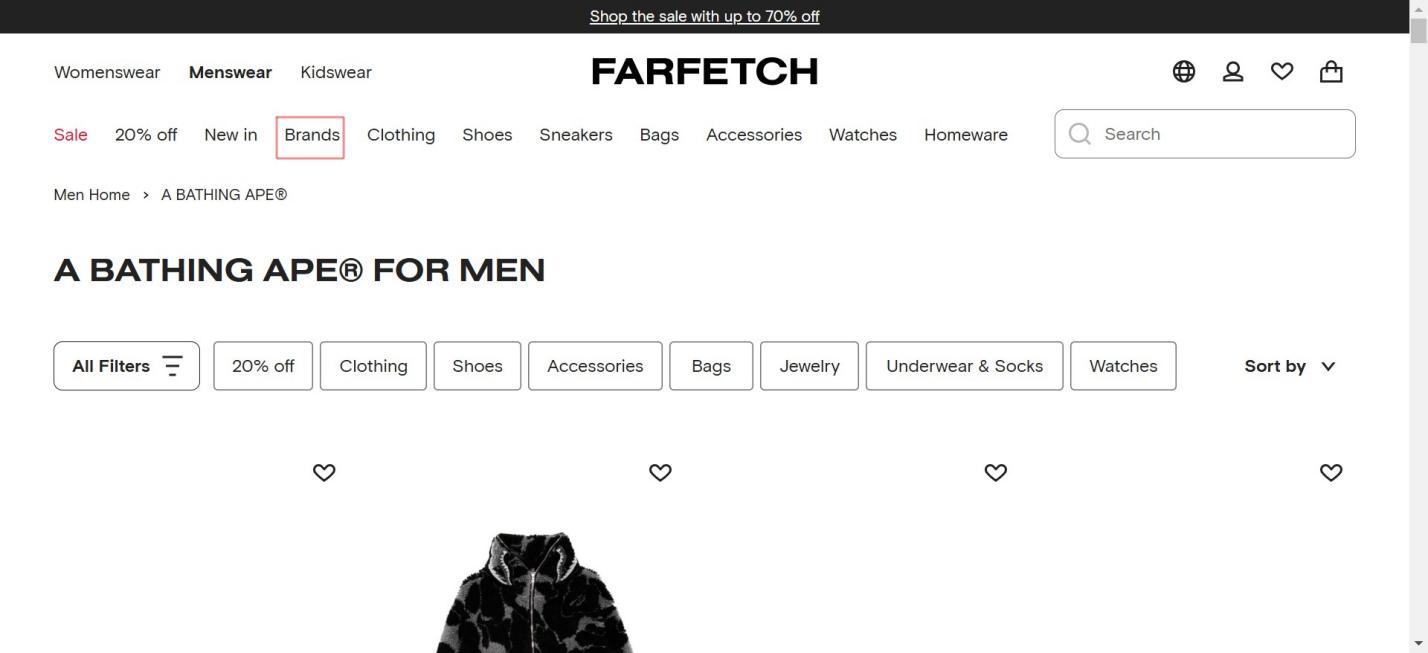
These are the fields that we need to grab.

Make **sure** that we can grab at least the **Title, Image urls, Product url, Availability, Price, currently and a Unique identifier.**

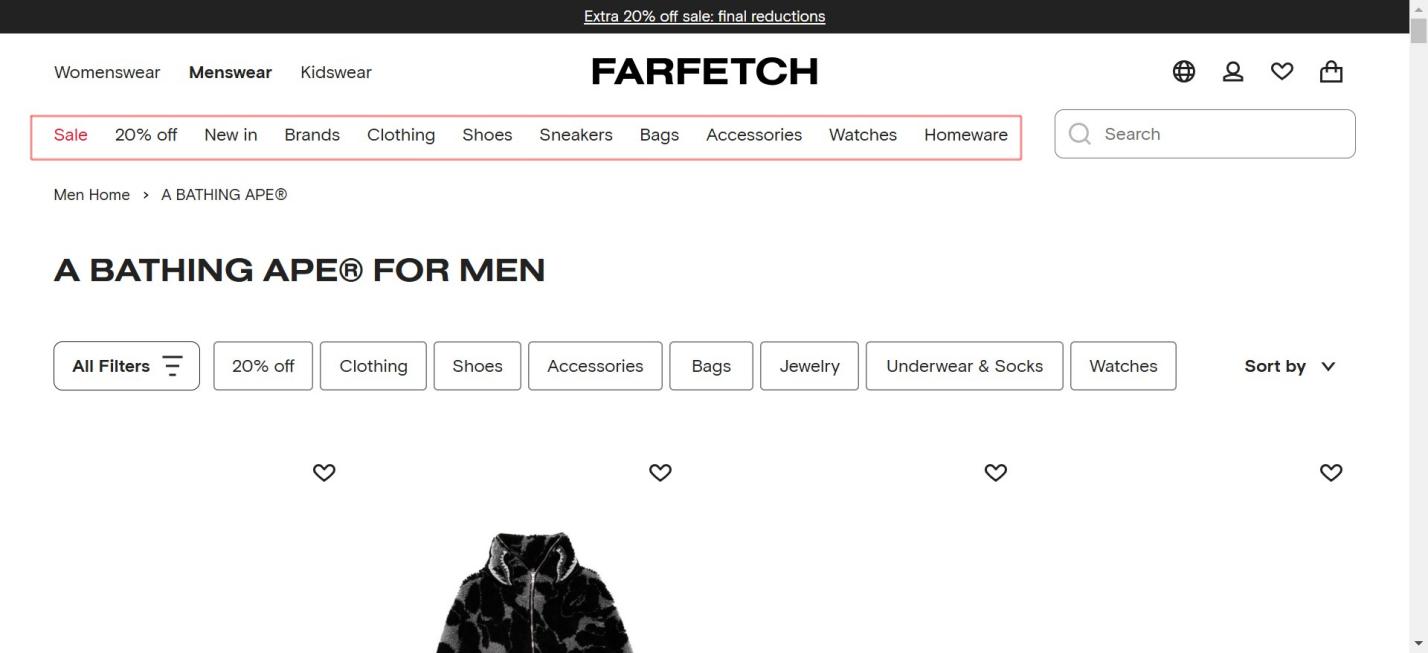
Three main Categories.



Brands are here against each main category.



And sub categories you can see here:



Rest of the data fields are on the detail page. And also scrape the rest of the data fields with a script.