

# **Product Materials & Price in the Fast Fashion Industry:**

## **Scraping the H&M website**

*Adapted from: Gebru, Morgenstern, Vecchione, Vaughan, Wallach,  
Daumeé, and Crawford. (2018). Datasheets for Datasets.*

<i>Fleur Le Mire</i>	2074117
<i>Mariëlla van Erve</i>	2035793
<i>Nishtha Stacie</i>	2008350
<i>Hilal Türer</i>	2080149
<i>Yi Ting Tsai</i>	2101518

# 1. Motivation

## 1.1 For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

The fashion industry is the third-largest manufacturing sector in the world (Papazolomou, 2022). In the early 1990s, fast fashion saw a surge of interest and quickly reached a significant level of popularity within a relatively short period of time. This trend towards fast fashion not only revolutionized the way clothing was produced and consumed, but also had significant economic and environmental impacts on the industry and society as a whole. The term “fast fashion” was first used at the beginning of the 1990s, when Zara opened in New York (Safdie & Gendre, 2022). Its’ strengths lie in the opportunity it provides to young consumers to follow fashion trends at an affordable price, and therefore has continued to skyrocket at the pace it already has. However, since fast fashion is inexpensive it promotes unnecessary, excessive consumption (Safdie & Gendre, 2022). As a result, it contributes a devastating impact on the planet (Segran, 2021; Brightly, 2022). Sustainable development and the environment suffer because of the rapid fashion business (Niinimäki, 2020). Furthermore, fast fashion makes use of unsustainable materials which further impact the environment negatively. In fact, one-third of microplastics in the ocean is from the materials used to produce fast fashion (Safdie & Gendre, 2022).

As consumers have become more aware of the negative impact that unsustainable product materials can have on the environment, there has been a growing trend towards sustainability and eco-friendliness in the fashion industry. In response, many fashion companies have started to invest in more sustainable and environmentally-friendly production methods and materials, in order to reduce their environmental footprint and meet the changing demands of their customers. Relatedly, many fast fashion companies have started to incorporate “green marketing” into their strategies, which was a tool to attract environmentally conscious customers (Mandaric et al., 2022).

For instance, H&M publicly set the goal of using only recycled and sustainably-sourced materials by 2030. These initiatives enable the company to set up a greener fashion image. According to its 2021 Sustainability Report, the brand has tripled the share of recycled materials used in its garments as well as reduced its plastic packaging by nearly 28% (Igini, 2022).

Despite the public statements of these biggest fast fashion retail companies, there are still some questions in mind. These questions are about their sustainability campaigns, and whether they are partaking in greenwashing practices. Greenwashing is a marketing strategy where companies present their products, services or brand in a way that makes consumers believe it is environmentally friendly (Igini, 2022).

Another dimension worth looking into is the prices of these fast fashion brands. The prices of sustainable fabrics are, on average, two and a half to four times more expensive in comparison to more commonly used fabrics (Wanderlust, 2021). Since fast fashion brands still operate at low and affordable prices, it raises the question about their sustainability.

These revelations have suggested that it is worth looking into the fast fashion industry. This study opted to focus on collecting data on the product materials & price of H&M products. By collecting this data researchers will be able to have information on one of the leading fast fashion brands. This data collection will focus on only women’s products since most consumers of fast fashion are women and generally, women’s fashion is a bigger market than men’s fashion (Cross, 2022). With the collected data, researchers can look into whether the products are made from sustainable materials or do research on the correlation between price and product material.

To collect this data, multiple data sources and extraction methods have been considered as shown in Appendix A1 (Data Source Comparison Table). Ten well-known fast fashion brands, like for example H&M, Zara, and Uniqlo, are browsed through to explore the universe of potential web sources. These fast fashion brands were obtained by using Google Trends and searching on Google for 'Fast fashion brands'. The metrics that can be scraped for each source and the conditions necessary to access the data are added to the table. This includes metrics like for example material information, the price, colors, and sizes.

There are alternative ways to generate data on this topic. The alternatives this study considered were publicly accessible APIs (either free or requiring a payment), or using an existing dataset on for example Kaggle. However, the majority of APIs that can be used require payment. Examples of APIs that could be used are the ShoppingScraper API of RapidAPI. Both of these require a payment. Because we do not know the quality of the API, we have not opted for this. In addition, the costs are high if this is used for a number of weeks and this is not affordable for us at this point in time.

This research aims to gain insight into the sustainability of the materials and the correlation with the price. That is why it is important that the website that is going to be scraped also contains the material information. H&M, Forever21, and New Look are the fast fashion brands, of which the website contains material information to scrape, and that do have overview pages with all products for women. Zara, Uniqlo, and Shein do also have material information on their products but do not have one overview page which contains all of the products. The Dutch website of H&M roughly has 17,140 women's products available. However, only 9,972 products are available to load with a scraper or manually on the webpage. Their URLs of the 'bekijk alle items' page are based on the number of products (e.g. a page contains 36 products). When you click on the button 'Meer artikelen laden', you will see 72 products in the URL and so further on. New Look has a clear product page URL, which goes up by 1 if you enter the next page. However, New Look only has 5,777 products. Forever21 has 8,156 products and works with similar URLs as H&M. Forever21 is less sustainable according to Good on You (Good on You, 2021). Therefore Forever21 is the least interesting compared to H&M and New Look.

H&M has begun broadcasting and focussing on more sustainable practices and has a lot of products to scrape compared to the other fast fashion brands. For that reason, this research project is focusing on H&M. [https://www2.hm.com/nl\\_nl/index.html](https://www2.hm.com/nl_nl/index.html) is the official website of H&M in the Netherlands and has been used for the research project. H&M's website can be considered a dynamic website since the product quantity and prices on the website fluctuate, such as during sales events. It is also crucial to include the date of scraping into the dataset. The researchers can also investigate the trend over time in the future.

## **1.2 Who created this dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?**

The dataset that is available after scraping the website [www2.hm.com](https://www2.hm.com/nl_nl/index.html) has been scraped by a project group (Team 8) of the course Online Data Collection and Management. This course has been given by Hannes Datta at Tilburg University and is part of the master Marketing Analytics.

## **1.3 Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.**

The creation of the dataset was not directly funded by any external partner. It has been made possible through the teaching materials provided by Tilburg University. These materials equipped the project group members with the necessary information to scrape the [www2.hm.com/nl\\_nl/](https://www2.hm.com/nl_nl/) website.

## 2. Composition

**2.1 What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.**

The instances in the dataset(s) are women's items extracted from the H&M website in The Netherlands. Each instance gives information of the title of the product, the price, the color, material information and if the product is a new arrival item. To extract the information of instances, the product overview pages which are the pages showing the total offer of products on a website, and product pages are both used as seeds for data extraction.

**2.2 How many instances are there in total (of each type, if appropriate)?**

The dataset contains 7,904 instances from "VIEW ALL ITEMS" of women's items on the H&M website. The dataset doesn't contain all 9,772 instances on the website. In section 2.3, we provide more detail on the sampling strategy we used for the data collection.

**2.3 Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).**

The dataset contains samples of instances from "VIEW ALL ITEMS" of women's items on the H&M website. There is an infinite loading button on the product overview page instead of a next page button. As a result, the page will aggregately load 36 more items whenever pressing the loading button. As of the extracting date, 17,140 products were shown on the product overview page. However, after iterating through their pages to explore the limits, there are solely 9,972 products that are loadable.

On the other hand, in the data collection process, not only the number of items that can be extracted but also the scraping time needs to be taken into account. By our calculations, it takes about 10 minutes to extract all product URLs from the overview page. In addition, it takes approximately 35 seconds to extract the product information from a product page, which leads to 97 hours in total to extract the product information from all of the product pages. In order to extract data within a technically feasible time frame and to consider the reasonable sample size, we decided to use `random.shuffle()` function to scrape data from four different devices for 48 hours. After scraping, we process to combine the data sets and remove duplicate instances. Based on the sampling strategy, we estimated to collect 75% out of 9,972 product pages, meaning 7,479 product pages of women's items on the H&M website we are aiming to collect. We deemed 75% as a suitable representation of the total population.

Finally, we have scraped 7,904 instances after removing duplicates which is around 79.3% of the total loadable product pages.

**2.4 What data does each instance consist of? "Raw" data (e.g., unprocessed text or images) or features? In either case, please provide a description.**

For each instance, a women's item, the following data is collected:

Variable	Description	Type of data
Title	The name of the product.	Unprocessed text
Price	The price of the product.	Unprocessed text
Color	The color of the product.	Unprocessed text
Product URL	The url of the product.	Unprocessed text
Buitenlaag	The outside material of the product.	Unprocessed text
Material	The material of the product.	Unprocessed text
New arrival	If the product is a new arrival item or a special label item	Unprocessed text
Timestamp	The date and time when scraping events have occurred.	Unprocessed text

Some products consist of several different materials and several parts, as a result, this information was extracted from distinct HTML elements. We separated them into two variables called “Buitenlaag” and “Material.”

## 2.5 Is there a label or target associated with each instance? If so, please provide a description.

There is no label or target created to be associated with each instance.

## 2.6 Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.

The variables that contain missing values are materials, buitenlaag, and new arrival. The *materials* variable contained missing values for products that did not show information for materials, these values show in the dataset as “no materials”. Moreover, for the *buitenlaag* variable the missing values were due to some products not having information available about the outer layer of the products. These missing values show in the dataset as “no buitenlaag”. Missing values result from the fact that some products only have one of these material characteristics while others possess both the buitenlaag and material information. Finally, for the *new arrival* variable, it shows a missing value when a product was not positioned on the website as a new arrival. These missing values were converted to “no new arrival” in the dataset.

## 2.7 Are relationships between individual instances made explicit (e.g., users’ movie ratings, social network links)? If so, please describe how these relationships are made explicit.

The explicit relationships between individual instances do not exist because the collected instances are independent.

## 2.8 Are there recommended data splits (e.g., training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.

The product information of each product is collected and saved in one dataset. In our case, we decided not to split the dataset right now. However, it is possible for researchers to split the dataset in the future by material, color or other relevant parameters, depending on the research objectives.

**2.9 Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a future user? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.**

The dataset is self-contained without relying on other external sources.

**2.10 Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor patient confidentiality, data that includes the content of individuals non-public communications)? If so, please provide a description.**

As mentioned in 2.4, all the data we scraped is the product information which is publicly available on the H&M website. There is no personal information scraped in this data set too.

**2.11 Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.**

The dataset does not contain any offensive, insulting, threatening or anxiety causing information.

**2.12 Does the dataset relate to people? If not, you may skip the remaining questions in this section.**

There is no user-related information extracted during the collection of the data.

**2.13 Does the dataset identify any subpopulations (e.g., by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.**

As mentioned in 2.12, the dataset doesn't include any user-related information during extraction. As a result, the dataset cannot identify any subpopulations.

**2.14 Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset? If so, please describe how.**

Since the data within the dataset are all related to product information, which is also publicly shown on the H&M website, there is no user information that can be identified from the dataset.

**2.15 Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals racial or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.**

The dataset doesn't contain data that might be considered sensitive.

### 3. Collection Process

**3.1 How was the data associated with each instance acquired? Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.**

The data is acquired by scraping H&M's women all items pages. All acquired data was directly observable on the H&M's website.

**3.2 What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)? How were these mechanisms or procedures validated?**

The mechanism used to collect the data from H&M's website is a semi-manually programmed web scraper. The web scraper is programmed in a Jupyter Notebook with Python language, and used a combination of Selenium and BeautifulSoup. BeautifulSoup has been used to collect the product URLs from the product overview pages. A code python with BeautifulSoup code has been used to counter the number of articles in the page URL and to make a list of the page URLs up to the last page of the 9,972 products. Meanwhile, Selenium is used to scrape the product information on product pages and open boxes of further information.

**3.3 If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?**

As far as [www2.hm.com/nl\\_nl](http://www2.hm.com/nl_nl) is concerned, we scraped approximately 79.3% of the women's items that were available from 20/03/2023 until 23/03/2023 from the H&M's 'view all' section for our dataset. Our reason to acquire this sample is that H&M has thousands of products and the most popular products being women's clothing. Regarding the data collection, H&M's "view all" section was scraped once because 7,904 items are enough to answer research questions on product material and/or price. Furthermore, it is unlikely that the products' composition of a particular material will change. If this does change, the existing product will probably be dropped and replaced by a new product. Lastly, there is little chance that the price of the product will change in a short time.

**3.4 Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?**

The data is entirely collected by a course project group from Tilburg University. Therefore, no one receives any financial compensation.

**3.5 Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.**

The dataset contains observations captured from the 20th of March 2023 until the 23rd of March 2023. The scraping has been done only once, considering that price and product material will not change in a short span of time.

Four different devices have been used during the scraping period to collect the data. The first device scraped for approximately 64 hours. The second device scraped for about 48 hours. The third and fourth devices scraped for a

shorter period of time. The third one had scraped product information for about 28 hours. Lastly, the fourth device had been used for scraping for approximately 10 hours. Each device could scrape about 100 products in an hour. However, due to the use of multiple devices, it was not possible to guarantee the exclusion of duplicates. At the end, there were roughly 5,000 scraped duplicates. When the data was cleaned, these duplicates were removed.

**3.6 Were any ethical review processes conducted (e.g., by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.**

For web scraping projects, it is important to ensure that all ethical principles and guidelines are adhered to avoid any potential legal or ethical issues. In the terms and conditions of the Dutch website, H&M prohibits using any engine, software, tool, agent, or other devices on their website or copying their information. In that respect, it is not ethical to scrape the website. However, since no agreement has been signed, H&M's publicly available product data can be scraped. There is no personal or sensitive information on the H&M website. Moreover, it only contains publicly available information. Therefore, this scraping project is not subject to ethical concerns since the only objective is to gain familiarity with scraping within the context of Online Data Collection & Management course at Tilburg University.

**3.7 Does the dataset relate to people? If not, you may skip the remaining questions in this section.**

Since the dataset solely contains product information as specified by H&M, it does not relate to individuals.

**3.8 Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)?**

Not applicable.

**3.9 Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how the notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.**

Not applicable.

**3.10 Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.**

Not applicable.

**3.11 If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).**

Not applicable.

**3.12 Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.**

Not applicable.



## 4. Preprocessing, cleaning, labeling

**4.1 Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remainder of the questions in this section.**

The dataset of women products' information from H&M is extracted during the data collection process for this project. Due to our sampling strategy which is mentioned in section 2.3, we have to combine several data sets from scraping. After combining and inspecting the data, we found some issues to address:

- Duplication: Since we used the `random.shuffle()` function from four devices to scrape, it is possible to have duplicated rows once the data sets are combined. In the end, approximately 42% of instances are duplicated so we used R to remove the duplicated instances from the dataset.
- Timestamp: The column of "timestamp" shows the unix timestamp from the scraped data. As a result, we created a new column called "date" to convert the unix timestamp to a normal date.
- Price: The column of "price" contained the metadata of the euro dollar sign which is not necessary to use in research. As a result, we removed it using R as well.

**4.2 Was the "raw" data saved in addition to the pre-processed/cleaned/labeled data (e.g., to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.**

The raw data was saved as a Json file. The file can be found on the following link:

<https://github.com/Mariellavanerve/odcm-project>

**4.3 Is the software used to preprocess/clean/label the instances available? If so, please provide a link or other access point.**

After scraping the data sets by Jupyter Notebook, we used R to conduct the data cleaning process. The source code can be found on the following link: <https://github.com/Mariellavanerve/odcm-project>

## 5. Uses

**5.1 Has the dataset been used for any tasks already? If so, please provide a description.**

This dataset has not been used for any tasks before, as this dataset is specially developed as a project for the course Online Data Collection and Management as part of the Master Marketing Analytics at Tilburg University.

**5.2 Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.**

There is a repository that gives access to all documents that are related to this project. The repository is publically available and can be found via the following link: <https://github.com/Mariellavanerve/odcm-project>

**5.3 What (other) tasks could the dataset be used for?**

This data can be combined with data from other fast fashion brands to enrich and disseminate the research more widely. Another example might be that the outputs from H&M data can influence fast fashion manufacturers' decisions about price and product material selection.

**5.4 Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a future user might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other undesirable harms (e.g., financial harms, legal risks) If so, please provide a description. Is there anything a future user could do to mitigate these undesirable harms?**

Since the price information can vary depending on the season, and the pricing data will only reflect the March 2023 economies of scale, researchers should consider this, and should not do pricing analysis solely based on the price data of H&M in future uses. Moreover, data has been collected specifically on days that there were no member sales. During these days, the website lists a 20% discount off on the prices of all products. It has been determined not to scrape on such days because doing so would lead to false conclusions about H&M's prices.

**5.5 Are there tasks for which the dataset should not be used? If so, please provide a description.**

This data has been collected to provide a source for research that would like to gather insights. Therefore, it should be only used for research purposes. The reason for this is that H&M prohibits using and copying its content from the website. When using it for commercial purposes, you could face legal action for that.

## References

- Brightly. (2022, August 2) The most popular fast fashion brands, ranked for conscious consumers. Brightly. Retrieved February 22, 2023, from <https://brightly.eco/blog/fast-fashion-brands-sustainability>
- Cross, R. (2022, September 7). Fast fashion for men: A lesser known evil. Impakter. Retrieved February 27, 2023, from <https://impakter.com/fast-fashion-for-men/#:~:text=Most%20consumers%20of%20fast%20fashion,bigger%20market%20than%20men%27s%20fashion>
- Good on You. (2022, February 2). How ethical is H&M? Retrieved March 23, from <https://goodonyou.eco/how-ethical-is-hm/>
- Good on You. (2021, November 3). How ethical is New Look? Retrieved March 23, from <https://goodonyou.eco/how-ethical-is-new-look/>
- Good on You. (2021, March 10). How ethical is Forever21? Retrieved March 23, from <https://goodonyou.eco/how-ethical-is-forever-21/>
- Igini, M. (2023, February 17). 5 fast-fashion brands called out for greenwashing. Earth.Org. Retrieved February 27, 2023, from <https://earth.org/fast-fashion-brands-greenwashing/>
- Mandaric, D., Hunjet, A., & Vuković, D. (2022). The Impact of Fashion Brand Sustainability on Consumer Purchasing Decisions. *Journal of Risk and Financial Management*, 15(4), 176. <https://doi.org/10.3390/jrfm15040176>
- Niinimäki, K., Peters, G., Bontempi, E., Perry, P., Rissanen, T., & Gwilt, A. (2020). The environmental price of fast fashion. *Nature Reviews Earth & Environment*, 1(4), 189–200. <https://doi.org/10.1038/s43017-020-0039-9>
- Papasolomou, I., Melanthiou, Y., & Tsamouridis, A. (2022). The fast fashion vs environment debate: Consumers' level of awareness, feelings, and behaviour towards sustainability within the fast-fashion sector. *Journal of Marketing Communications*, 1–19. <https://doi.org/10.1080/13527266.2022.2154059>
- Safdie, S., & Gendre, I. (2022). All you need to know about fast fashion. Greenly. Retrieved February 27, 2023, from <https://greenly.earth/en-us/blog/ecology-news/all-you-need-to-know-about-fast-fashion>
- Wanderlust. (2021, December 20). Looking at true cost of sustainable clothing. Wanderlust. Retrieved February 27, 2023, from <https://wanderlust.com/journal/looking-at-true-cost-of-sustainable-clothing/#:~:text=The%20fabrics%20are%20more%20expensive&text=%E2%80%9CThe%20prices%20of%20sustainable%20fabrics,to%20more%20commonly%20used%20fabrics.>

## Appendix

### A.1 Data Source Comparison Table

Brand	Source	Description	Information	Page available with all products within one category	Web scraping or API	Free or Paid access	Sign in?
H&M	<a href="https://www2.hm.com/nl/index.html">https://www2.hm.com/nl/index.html</a>	Official website of H&M, with all the product pages and product information which is available about all of the products. Possible to scrape.	<ul style="list-style-type: none"> <li>• Price</li> <li>• Reviews</li> <li>• Materials</li> <li>• Size</li> <li>• Length</li> <li>• Style</li> <li>• Description</li> <li>• Color</li> <li>• New arrival products</li> </ul>	Yes → dames, shop op item, bekijk alle items	Web Scraping or API	Free	No
Uniqlo	<a href="https://www.uniqlo.com/eu/en/home">https://www.uniqlo.com/eu/en/home</a>	Official website of Uniqlo Co., Ltd. It is a Japanese casual wear designer, manufacturer and retailer. with all the product pages and product information which is available about all of the products. Possible to scrape.	<ul style="list-style-type: none"> <li>• Size</li> <li>• Color</li> <li>• Price</li> <li>• Delivery</li> <li>• Material</li> <li>• Care instructions</li> <li>• Description</li> </ul>	No, all categories separate	Web scraping (also via mydataprovider)	Free	No
Shein	<a href="https://eur.shein.com/">https://eur.shein.com/</a>	Official website of Shein. Bad news about their bad working conditions. Shein has a lot of pop-ups on the website, which makes the scraping more hard.	<ul style="list-style-type: none"> <li>• Color</li> <li>• Price</li> <li>• Size</li> <li>• Length</li> <li>• Fit</li> <li>• Material</li> <li>• Details about fit</li> <li>• Reviews</li> </ul>	No, all categories separate	Web scraping or API (paid)	Free	No
Zara	<a href="https://www.zara.com/nl/">https://www.zara.com/nl/</a>	Official website of Zara. ZARA is a Spanish multinational retail clothing chain. It specializes in fast fashion, and sells clothing, accessories, shoes, beauty products and perfumes.	<ul style="list-style-type: none"> <li>• Sizes</li> <li>• Price</li> <li>• Description</li> <li>• Material</li> </ul>	No, all categories separate	Web scraping or API (paid)	Free	No

<b>Forever 21</b>	<a href="https://www.forever21.com">https://www.forever21.com</a>	Official American website of Forever 21. Forever 21 is a multinational fast fashion retailer headquartered in Los Angeles, California, United States. Multiple pop-ups on the website when you enter.	<ul style="list-style-type: none"> <li>• Price</li> <li>• Sizes</li> <li>• Color</li> <li>• Fit</li> <li>• Material</li> <li>• Care</li> <li>• Description</li> </ul>	Yes → women, view all	Web scraping of API (paid)	Free	No
<b>Primark</b>	<a href="https://www.primark.com/">https://www.primark.com/</a>	Primark Stores Limited is an Irish multinational fast fashion retailer with headquarters in Dublin, Ireland. It has stores across Europe and in the United States. Not that many products, only 631 for women. When clicking on some products, you get an error. No material information.	<ul style="list-style-type: none"> <li>• Price</li> <li>• Productcode</li> </ul>	Yes → women, bekijk alle kleding	Web scraping	Free	No
<b>New Look</b>	<a href="https://www.newlook.com">https://www.newlook.com</a>	Official website of New Look. New Look is a British global fashion retailer with a chain of high street shops.	<ul style="list-style-type: none"> <li>• Price</li> <li>• Sizes</li> <li>• Material</li> <li>• Care</li> <li>• Fit</li> <li>• Description</li> </ul>	Yes → womens view all	Web scraping	Free	No
<b>GAP</b>	<a href="https://www.gap.com/">https://www.gap.com/</a>	Official website of GAP. The Gap, Inc., commonly known as Gap Inc. or Gap, is an American worldwide clothing and accessories retailer. American website. Material not clear for every product.	<ul style="list-style-type: none"> <li>• Price</li> <li>• Color</li> <li>• Sizes</li> <li>• Description</li> <li>• Care</li> <li>• Fit</li> <li>• Material sometimes</li> </ul>	Yes → women, shop all styles	Web scraping	Free	No
<b>Esprit</b>	<a href="https://www.esprit.eu/">https://www.esprit.eu/</a>	Official website of Esprit. Esprit Holdings Limited is a publicly owned manufacturer of clothing, footwear, accessories, jewellery and housewares under the Esprit label. The company is headquartered in North Point, Hong Kong, and Ratingen, Germany.	<ul style="list-style-type: none"> <li>• Sizes</li> <li>• Color</li> <li>• Price</li> <li>• Material</li> <li>• Care</li> <li>• Fit</li> <li>• Delivery</li> </ul>	No, separate categories	Web scraping	Free	No

---

<b>Fashion Nova</b>	<a href="https://www.fashionnova.com/">https://www.fashionnova.com/</a>	Official website of Fashion Nova. Fashion Nova is an American fast fashion retail company. The company primarily operates online, but it also has five brick-and-mortar locations. Fashion Nova is known to use social media marketing, particularly on Instagram	<ul style="list-style-type: none"> <li>• Sizes</li> <li>• Price</li> <li>• Reviews</li> <li>• Material</li> <li>• Details</li> </ul>	No, separate categories	Web scraping	Free	No
---------------------	---	---	--	-------------------------	--------------	------	----