

Web Scraping Test Description

Objective:

The objective of this test is to assess your ability to scrape product data from a website effectively. You will be required to extract information for products from the website. The extracted data should include the product name, brand/manufacturer, price, and EAN code.

Instructions:

1. Website:

<https://shop.rewe.de/>

Or

<https://www.citygross.se/matvaror>

2. Requirements:

- Extract data for as many products as you can, ideally all.
- For each product, gather the following information:
 - Product name
 - Brand/Manufacturer
 - Price
 - EAN/GTIN (European Article Number) code – This one is mandatory
- Tools: You are free to use any web scraping tools or libraries of your choice (e.g., BeautifulSoup, Scrapy, Selenium, etc.). Ensure that the chosen tools are capable of extracting the required information accurately.

3. Submission:

- Once you have completed the scraping process, submit your code on GITHUB along with a brief explanation of your approach. The code should be well-commented and organized for clarity.

4. Assessment Criteria:

- Accuracy of the extracted data.
- Efficiency of the scraping process.
- Quality of code structure and documentation.

5. Screenshots:

- To understand what fields we need for the scrapes, we provided two screenshots with highlighted information that is needed. EANs/GTINs are not visible on the website but they can be found somewhere inside the structure of the website.

Screenshot 1: This screenshot illustrates the layout of the product page for Rewe.de website. In top right corner are the categories.

Shop.rewe.de

Sortiment ▾

Meine Produkte

Angebote

Mengenrabatt

Bio


Regional

Neu

Weinfreunde

International

Italienisch



REWE Bio Mozzarella 125g

Bio Mozzarella, 45% Fett i. Tr.

Marke: REWE Bio

NUTRI-SCORE **A B C D E** + GTIN/EAN

1,19 €

125g (1 kg = 9,52 €)

1 °P pro 2 Euro Warenwert ⓘ

Produktbildhinweis:

Der Nutri-Score-Wert auf dem Produktbild ist beispielhaft und kann vom tatsächlichen Wert abweichen.

Screenshot 2: This screenshot displays the layout of the product page for Citygross.se website. Use this screenshot as a reference to locate the necessary information for extraction.

CITYGROSS.SE:

CITY GROSS

Groceries Prescription Catering Offers Shops

The pantry +

Dairy, cheese & eggs +

Frozen +

Bread & bakery +

Candy +

Drink -

Non-alcoholic wine

Cider

Drink mixes & essences

Fruit drink & still drinks

Juice, fruit drinks & smoothies +

Coffee, tea & chocolate +

Soda +

Juice +

Sports & energy drink +

Water

Words, price SEK 105,00 + deposit

Compare price SEK 10,60/l

Offer applies to 2 purchases per customer

29⁹⁵ / pes

CT. price SEK 17,11/l

Select store

74⁹⁵ + deposit

Compare price SEK 12,49/l

Select store

27 + deposit

Words, price SEK 35,95 + deposit

Compare price SEK 33,28/l

Select store

13⁹⁵ / pes

CT. price SEK 0,70/pc

Select store

Multivitamin Drink 30%

GLOCK GOLD, 2l

19⁹⁵ / pes

Compare price SEK 9,98/l

Select store

Apple juice

BRAVO, 2l

+ GTIN/EAN

38⁹⁵ / pes

Compare price SEK 19,48/l

Select store

Coca-Cola Zero

COCA-COLA ZERO, 1,5l

20⁹⁵ + deposit

Compare price SEK 13,97/l

Select store

Coca-Cola

COCA-COLA, 1,5l

20⁹⁵ + deposit

Compare price SEK 13,97/l

Select store

Brewed coffee Medium roast

LÖFBERGS, 450g

46⁹⁵ / pes

Compare price SEK 104,33/kg

Select store