

Web Scrapping Requirements Document

Overview:

Extract product data from a list of approximately 474 public Zazzle product URLs. These pages may trigger CAPTCHA and involve dynamic content loading.

Input:

- A CSV file named `Cleaned_Zazzle_Product_Links.csv` containing ~474 product page URLs.
 - A screenshot has been provided showing the page layout and where each required field appears.
-

Output:

A single CSV file containing one row per product, with the following columns:

Column Name	Description
Product Title	The name of the product as it appears on the product page.
Design Description	The user-provided description of the product design.
Breadcrumb Category	The full breadcrumb path of the product category.
Product ID	Unique product ID, found in the right-hand info box.
Created Date	The date the product was created, found in the right-hand info box.
Tags / Keywords	Comma-separated list of tags associated with the product.
Price	Displayed product price in USD (e.g. "\$24.95").
Image URL	Direct URL to the main product image (not a thumbnail).

Notes:

- Some content (e.g., price, tags) may be loaded dynamically. The scraper should wait for full page rendering.
 - Ensure the CSV is saved with UTF-8 encoding.
 - All price values should include the dollar sign and match the on-page format.
-

Sample Output Format (CSV):

CSV

CopyEdit

Product Title,Design Description,Breadcrumb Category,Product ID,Created Date,Tags / Keywords,Price,Image URL

"Funny Cat Mug","A cute mug with a grumpy cat","Home > Kitchen > Mugs","1234567890","2023-08-14","cat, mug, funny, gift","\$14.95","https://rlv.zcache.com/image_url.jpg"

