



```
1  # =====
2  # IMPORTS
3  # =====
4  # Import required libraries for web scraping, browser automation, and data handling
5  import requests
6  from bs4 import BeautifulSoup
7  from playwright.sync_api import sync_playwright
8  import pandas as pd
9  import time
10 import os
11
12
13 def scrape_nawy_properties(
14     url="https://www.nawy.com/search?page_number=1&category=property",
15     scroll_count=100,
16     scroll_wait=5,
17     scroll_distance=1500,
18     container_selector="div.sc-88b4dfdb-0.cgVQXi",
19     property_class="sc-100c08da-0 eeBcMz",
20     output_path="real_estate_properties.csv"
21 ):
22     """
23     Scrapes real estate property data from Nawy.com using Playwright and BeautifulSoup.
24
25     Parameters:
26     - url (str): The URL of the Nawy search page to scrape.
27     - scroll_count (int): Number of times to scroll down to load more properties.
28     - scroll_wait (float): Time to wait (in seconds) between each scroll.
29     - scroll_distance (int): Pixels to scroll down per iteration.
30     - container_selector (str): CSS selector for the scrollable container.
31     - property_class (str): Class name of individual property listing elements.
32     - output_path (str): File path to save the resulting CSV.
33
34     Returns:
35     - pd.DataFrame: DataFrame containing scraped property data.
36     """
37
38     # =====
39     # INITIAL SETUP
40     # =====
41     # Make an initial GET request (kept for reference, not used in final scraping)
42     response = requests.get(url)
43
44
45     # =====
46     # LAUNCH BROWSER WITH PLAYWRIGHT
47     # =====
48     with sync_playwright() as p:
49         # Launch the browser in non-headless mode (visible window)
50         browser = p.chromium.launch(headless=True)
51
52         # Open a new browser page
53         page = browser.new_page()
54
55         # Navigate to the target URL
56         page.goto(url)
57
58         # Wait until the scrollable container (which holds the property listings) is loaded
59         page.wait_for_selector(container_selector)
60
61         # =====
62         # INFINITE SCROLL SIMULATION
63         # =====
64         # Scroll down inside the scrollable container multiple times to load all properties
65         for _ in range(scroll_count):
66             page.evaluate(f"""
67                 () => {{
68                     const container = document.querySelector('{container_selector}');
69                     if (container) {{
70                         container.scrollBy(0, {scroll_distance}); // Scroll down by specified pixels
71                     }}
72                 }}
73             """)
74             time.sleep(scroll_wait) # Wait for content to load
75
76         # =====
77         # EXTRACT FULL PAGE HTML AFTER SCROLLING
78         # =====
79         # After scrolling, retrieve the complete page HTML (now includes dynamically loaded content)
80         html = page.content()
81
82         # Parse the HTML using BeautifulSoup for easier data extraction
83         soup = BeautifulSoup(html, "html.parser")
84
85         # Find all property listing elements using the specified class name
86         Properties = soup.find_all('div', class_=property_class)
87
88
89         # =====
90         # DATA STORAGE INITIALIZATION
91         # =====
92         # Initialize empty lists to store extracted property data
93         location_list = []
94         name_list = []
95         description_list = []
96         area_list = []
97         bed_list = []
98         bath_list = []
99         price_list = []
100
101
102         # =====
103         # EXTRACT DATA FROM EACH PROPERTY
104         # =====
105         # Loop through each property element on the page
106         for property in Properties:
107             # Extract basic textual information using CSS selectors
108             location = property.select_one('div.area') # Location of the property
109             name = property.select_one('div.name') # Name/title of the property
110             description = property.select_one('h2.sc-4b9910fd-0.hyACaB') # Description headline
111             price = property.select_one('div.price-container span.price') # Price of the property
112
113             # Append text content if element exists; otherwise, append empty string
114             location_list.append(location.text.strip() if location else "")
115             name_list.append(name.text.strip() if name else "")
116             description_list.append(description.text.strip() if description else "")
117             price_list.append(price.text.strip() if price else "")
118
119             # Initialize default values for area, beds, and baths
120             area_val = ""
121             bed_val = ""
122             bath_val = ""
123
124             # =====
125             # EXTRACT FEATURE BLOCKS (AREA, BEDS, BATHS)
126             # =====
127             # Some properties display additional details in labeled feature blocks
128             feature_blocks = property.select("div.sc-234f71bd-0.bbWDeD") # Select all feature blocks
129
130             # Loop through each feature block (e.g., "m2", "beds", "baths")
131             for block in feature_blocks:
132                 label = block.select_one("span.label") # Label (e.g., "m2", "beds")
133                 value = block.select_one("span.value") # Value (e.g., "120", "3")
134
135                 # If both label and value exist, process them
136                 if label and value:
137                     label_text = label.text.strip().lower() # Normalize label to lowercase
138                     value_text = value.text.strip()
139
140                     # Match label to appropriate field and assign value
141                     if label_text == "m2":
142                         area_val = value_text
143                     elif label_text == "beds":
144                         bed_val = value_text
145                     elif label_text == "baths":
146                         bath_val = value_text
147
148             # Append extracted feature values to their respective lists
149             area_list.append(area_val)
150             bed_list.append(bed_val)
151             bath_list.append(bath_val)
152
153         # =====
154         # CLOSE BROWSER
155         # =====
156         # Close the browser after scraping is complete
157         browser.close()
158
159
160         # =====
161         # CREATE PANDAS DATAFRAME
162         # =====
163         # Combine all data lists into a structured DataFrame
164         df = pd.DataFrame({
165             'Location': location_list,
166             'Name': name_list,
167             'Description': description_list,
168             'Area': area_list,
169             'Beds': bed_list,
170             'Baths': bath_list,
171             'Price': price_list
172         })
173
174
175         # =====
176         # DEBUG OUTPUT: PRINT SCRAPED DATA
177         # =====
178         # Print all collected data to verify successful scraping
179         print("Locations:", location_list)
180         print("Names:", name_list)
181         print("Descriptions:", description_list)
182         print("Prices:", price_list)
183         print("Areas:", area_list)
184         print("Beds:", bed_list)
185         print("Baths:", bath_list)
186
187
188         # =====
189         # DISPLAY DATAFRAME INFO
190         # =====
191         # Print basic information about the scraped
```