

```

1  # -*- coding: cp1251 -*-
2  import csv
3  import os
4  import threading
5  import time
6  import winsound
7  from multiprocessing import Pool
8
9  import selenium.webdriver.support.expected_conditions as EC
10 import undetected_chromedriver as uc
11 from selenium.webdriver.common.action_chains import ActionChains
12 from selenium.webdriver.remote.webdriver import By
13 from selenium.webdriver.support.wait import WebDriverWait
14
15 my_path = f"{os.getcwd()}\\CIAN_diplom"
16
17 with open(rf"{my_path}\\df_flats.csv", "a", encoding="utf-8", newline="") as ouf:
18     writer = csv.writer(ouf, delimiter="$", quotechar="|", quoting=csv.QUOTE_MINIMAL)
19     writer.writerow(
20         (
21             "link",
22             "last_update",
23             "title",
24             "JK",
25             "start_price",
26             "start_date",
27             "building",
28             "date_readiness",
29             "full_price",
30             "price_for_sq_meter",
31             "metro_station",
32             "metro_distance",
33             "city",
34             "district",
35             "area",
36             "street",
37             "house",
38             "full_square",
39             "living_square",
40             "kitchen_square",
41             "floor",
42             "finishing",
43             "developer",
44             "date_creation",
45             "info_about_flat_1",
46             "info_about_flat_2",
47             "info_about_flat_3",
48             "info_about_flat_4",
49             "info_about_flat_5",
50             "info_about_flat_6",
51             "info_about_flat_7",
52             "info_about_flat_8",
53             "info_about_flat_9",
54             "info_about_house_1",
55             "info_about_house_2",
56             "info_about_house_3",
57         )
58     )
59
60 xpath_dict = {
61     "from_date": '//*[@id="frontend-offer-card"]/div[1]/div[1]/div[2]/div[1]/div[1]/span',
62     "from_date2": '//*[@id="frontend-offer-card"]/main/div[1]/div[2]/div[1]/div[1]',
63     "title": '//*[@id="frontend-offer-card"]/div[1]/div[2]/section/div/div/div[1]/h1',
64     "title2": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/section/div/div[1]/div[1]/h1',
65     "JK": '//*[@id="frontend-offer-card"]/div[1]/div[2]/section/div/div/div[1]/div/a',
66     "JK2": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/section/div/div[1]/div[2]/a',
67     "building": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/section/div/div[1]/div[2]/span[1]',
68     "building2": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/section/div/div/div[2]/span[1]',
69     "date_readiness": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/section/div/div[1]/div[2]/span[2]',
70     "date_readiness2": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/section/div/div[1]/div[2]/span',
71     "full_price": '//*[@id="frontend-offer-card"]/div[1]/div[3]/div/div[1]/div[1]/div[1]/div[3]/div/div[1]/span',
72     "full_price2": '//*[@id="frontend-offer-card"]/main/div[3]/div/div[1]/div[1]/div[2]/div/div[1]/div/span/span[1]',
73     "price_sq": '//*[@id="frontend-offer-card"]/div[1]/div[3]/div/div[1]/div[1]/div[3]/div/div/div[1]/p[2]',
74     "price_sq2": '//*[@id="frontend-offer-card"]/main/div[3]/div/div[1]/div[1]/div[2]/div/div[1]/div/div[1]/span',
75     "metro_station": '//*[@id="frontend-offer-card"]/div[1]/div[2]/section/div/div/div[2]/address/ul[1]/li[1]/a',
76     "metro_station2": '//*[@id="frontend-offer-card"]/div[1]/div[2]/section/div/div/div[2]/address/ul[2]/li[1]/a',
77     "metro_station3": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/section/div/div[1]/div[3]/ul[1]/li[1]/a',
78     "metro_station4": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/section/div/div[1]/div[3]/ul[2]/li[1]/a',
79     "metro_distance": '//*[@id="frontend-offer-card"]/div[1]/div[2]/section/div/div/div[2]/address/ul[1]/li[1]/span',
80     "metro_distance2": '//*[@id="frontend-offer-card"]/div[1]/div[2]/section/div/div/div[2]/address/ul[2]/li[1]/span',
81     "metro_distance3": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/section/div/div[1]/div[3]/ul[1]/li[1]/span',
82     "metro_distance4": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/section/div/div[1]/div[3]/ul[2]/li[1]/span/span',
83     "city": '//*[@id="frontend-offer-card"]/div[1]/div[2]/section/div/div/div[2]/address/div/div/a[1]',
84     "city2": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/section/div/div[1]/div[3]/address/a[1]',
85     "district": '//*[@id="frontend-offer-card"]/div[1]/div[2]/section/div/div/div[2]/address/div/div/a[2]',
86     "district2": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/section/div/div[1]/div[3]/address/a[2]',
87     "area": '//*[@id="frontend-offer-card"]/div[1]/div[2]/section/div/div/div[2]/address/div/div/a[3]',
88     "area2": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/section/div/div[1]/div[3]/address/a[3]',
89     "street": '//*[@id="frontend-offer-card"]/div[1]/div[2]/section/div/div/div[2]/address/div/div/a[4]',
90     "street2": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/section/div/div[1]/div[3]/address/a[4]',
91     "house": '//*[@id="frontend-offer-card"]/div[1]/div[2]/section/div/div/div[2]/address/div/div/a[5]',
92     "house2": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/section/div/div[1]/div[3]/address/a[5]',
93     "full_square": '//*[@id="frontend-offer-card"]/div[1]/div[2]/div[3]/div[1]/div[2]/span[2]',
94     "full_square2": '//*[@id="frontend-offer-card"]/div[1]/div[2]/div[4]/div[1]/div[2]/span[2]',
95     "full_square3": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/div[2]/div/div[1]/div[1]',
96     "living_square": '//*[@id="frontend-offer-card"]/div[1]/div[2]/div[3]/div[2]/div[2]/span[2]',
97     "living_square2": '//*[@id="frontend-offer-card"]/div[1]/div[2]/div[4]/div[2]/div[2]/span[2]',
98     "living_square3": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/div[2]/div/div[2]/div[1]',
99     "kitchen_square": '//*[@id="frontend-offer-card"]/div[1]/div[2]/div[3]/div[3]/div[2]/span[2]',
100    "kitchen_square2": '//*[@id="frontend-offer-card"]/div[1]/div[2]/div[4]/div[3]/div[2]/span[2]',
101    "kitchen_square3": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/div[2]/div/div[3]/div[1]',
102    "floor": '//*[@id="frontend-offer-card"]/div[1]/div[2]/div[3]/div[4]/div[2]/span[2]',
103    "floor2": '//*[@id="frontend-offer-card"]/div[1]/div[2]/div[4]/div[4]/div[2]/span[2]',
104    "floor3": '//*[@id="frontend-offer-card"]/main/div[2]/div[1]/div[2]/div/div[4]/div[1]',
105    "finishing": '//*[@id="frontend-offer-card"]/div[1]/div[2]/div[3]/div[7]/div[2]/span[2]',
106    "finishing2": '//*[@id="frontend-offer-card"]/div[1]/div[2]/div[3]/div[6]/div[2]/span[2]',
107    "finishing3": '//*[@id="frontend-offer-card"]/div[1]/div[2]/div[3]/div[6]/div[2]/span[2]',

```

```

107 "finishing3": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[4]/div[1]/div[2]/span[2]',
108 "finishing4": '/*[@id="frontend-offer-card"]/main/div[2]/div[1]/div[2]/div/div[5]/div[1]',
109 "developer": '/*[@id="author"]/div[2]/a/h2',
110 "developer2": '/*[@id="author"]/div/div/div/div[1]/div[1]/span[2]',
111 "developer3": '/*[@id="author"]/div[2]/span',
112 "button_date_creation": '/*[@id="frontend-offer-card"]/div[1]/div[1]/div[2]/div[1]/div[2]/button[@data-name="OfferStats"]',
113 "button_line_creation2": '/*[@id="frontend-offer-card"]/main/div[1]/div[2]/div[1]/div[2]/a',
114 "date_creation": "/html/body/div[5]/div/div/div[2]/div/div/div/div[1]/div[1]",
115 "info about flat_1": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[5]/div/div/div/div[1]/div[1]/div[2]/div[1]/p[2]',
116 "info about flat_2": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[5]/div/div/div/div[1]/div[1]/div[2]/div[2]/p[2]',
117 "info about flat_3": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[5]/div/div/div/div[1]/div[1]/div[2]/div[3]/p[2]',
118 "info about flat_4": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[5]/div/div/div/div[1]/div[1]/div[2]/div[4]/p[2]',
119 "info about flat_5": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[5]/div/div/div/div[1]/div[1]/div[2]/div[5]/p[2]',
120 "info about flat_6": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[5]/div/div/div/div[1]/div[1]/div[2]/div[6]/p[2]',
121 "info about flat_7": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[5]/div/div/div/div[1]/div[1]/div[2]/div[7]/p[2]',
122 "info about flat_8": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[5]/div/div/div/div[1]/div[1]/div[2]/div[8]/p[2]',
123 "info about flat_9": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[5]/div/div/div/div[1]/div[1]/div[2]/div[9]/p[2]',
124 "info about flat_11": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[6]/div/div/div/div[1]/div[1]/div[2]/div[1]/p[2]',
125 "info about flat_21": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[6]/div/div/div/div[1]/div[1]/div[2]/div[2]/p[2]',
126 "info about flat_31": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[6]/div/div/div/div[1]/div[1]/div[2]/div[3]/p[2]',
127 "info about flat_41": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[6]/div/div/div/div[1]/div[1]/div[2]/div[4]/p[2]',
128 "info about flat_51": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[6]/div/div/div/div[1]/div[1]/div[2]/div[5]/p[2]',
129 "info about flat_61": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[6]/div/div/div/div[1]/div[1]/div[2]/div[6]/p[2]',
130 "info about flat_71": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[6]/div/div/div/div[1]/div[1]/div[2]/div[7]/p[2]',
131 "info about flat_81": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[6]/div/div/div/div[1]/div[1]/div[2]/div[8]/p[2]',
132 "info about flat_91": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[6]/div/div/div/div[1]/div[1]/div[2]/div[9]/p[2]',
133 "info about flat_12": '/*[@id="frontend-offer-card"]/main/div[2]/div[3]/div[2]/div/div/section/article/ul/li[1]/span[2]',
134 "info about flat_22": '/*[@id="frontend-offer-card"]/main/div[2]/div[3]/div[2]/div/div/section/article/ul/li[2]/span[2]',
135 "info about flat_32": '/*[@id="frontend-offer-card"]/main/div[2]/div[3]/div[2]/div/div/section/article/ul/li[3]/span[2]',
136 "info about flat_42": '/*[@id="frontend-offer-card"]/main/div[2]/div[3]/div[2]/div/div/section/article/ul/li[4]/span[2]',
137 "info about flat_52": '/*[@id="frontend-offer-card"]/main/div[2]/div[3]/div[2]/div/div/section/article/ul/li[5]/span[2]',
138 "info about flat_62": '/*[@id="frontend-offer-card"]/main/div[2]/div[3]/div[2]/div/div/section/article/ul/li[6]/span[2]',
139 "info about flat_72": '/*[@id="frontend-offer-card"]/main/div[2]/div[3]/div[2]/div/div/section/article/ul/li[7]/span[2]',
140 "info about flat_82": '/*[@id="frontend-offer-card"]/main/div[2]/div[3]/div[2]/div/div/section/article/ul/li[8]/span[2]',
141 "info about flat_92": '/*[@id="frontend-offer-card"]/main/div[2]/div[3]/div[2]/div/div/section/article/ul/li[9]/span[2]',
142 "info about house_1": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[5]/div/div/div/div[1]/div[2]/div[2]/div[1]/p[2]',
143 "info about house_2": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[5]/div/div/div/div[1]/div[2]/div[2]/div[2]/p[2]',
144 "info about house_3": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[5]/div/div/div/div[1]/div[2]/div[2]/div[3]/p[2]',
145 "info about house_11": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[6]/div/div/div/div[1]/div[2]/div[2]/div[1]/p[2]',
146 "info about house_21": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[6]/div/div/div/div[1]/div[2]/div[2]/div[2]/p[2]',
147 "info about house_31": '/*[@id="frontend-offer-card"]/div[1]/div[2]/div[6]/div/div/div/div[1]/div[2]/div[2]/div[3]/p[2]',
148 "start price button": '/*[@id="frontend-offer-card"]/div[1]/div[3]/div/div[1]/div[1]/div[1]/div[3]/div/div[2]/div/div/button[@data-name="Price',
149 "start price button2": '/*[@id="frontend-offer-card"]/main/div[3]/div/div[1]/div/div[1]/div/div[1]/div[class="a10a3f92e9--icon--Ri',
150 "start price": '/*[@id="frontend-offer-card"]/div[1]/div[3]/div/div[1]/div[1]/div[1]/div[3]/div/div[2]/div/div/div/div/div/div/div/div[3]/tabl',
151 "start price2": '/*[@id="frontend-offer-card"]/main/div[3]/div/div[1]/div[1]/div[2]/div/div[1]/div/div[1]/div/div/div/div/div/div/div/div[3]/table',
152 "start date": '/*[@id="frontend-offer-card"]/div[1]/div[3]/div/div[1]/div[1]/div[1]/div[3]/div/div[2]/div/div/div/div/div/div/div/div[3]/table',
153 "start date2": '/*[@id="frontend-offer-card"]/main/div[3]/div/div[1]/div[position()<2]/div[2]/div/div[1]/div/div[1]/div/div/div/div/div/div/div/div',
154 "unpublished": '/*[@id="frontend-offer-card"]/div[1]/div[1]/div[@data-name="OfferUnpublished"]',
155 "unpublished2": '/*[@id="frontend-offer-card"]/main/div[1]/div[@data-name="OfferUnpublished"]',
156 "captcha": '/*[@id="captcha"]',
157 }
158
159 url_list = []
160
161 with open(rf"{my_path}\flat_urls.txt", "r", encoding="utf-8") as f:
162     for line in f:
163         url_list.append(line.rstrip())
164
165 whole_time = time.time()
166
167 threadLocal = threading.local()
168
169
170 def get_driver() -> None:
171     """
172     Create webdriver and reassign attributes to use same driver multiple times
173     """
174     driver = getattr(threadLocal, "driver", None)
175     chromeOptions = uc.ChromeOptions()
176     chromeOptions.add_argument("--headless")
177     if driver is None:
178         driver = uc.Chrome(chrome_options=chromeOptions)
179     setattr(threadLocal, "driver", driver)
180     return driver
181
182
183 def melody() -> None:
184     """
185     Signals to notify about captcha
186     """
187     winsound.PlaySound("SystemExit", winsound.SND_ALIASES)
188     winsound.PlaySound("SystemExit", winsound.SND_ALIASES)
189     winsound.PlaySound("SystemHand", winsound.SND_ALIASES)
190     winsound.PlaySound("SystemExit", winsound.SND_ALIASES)
191     winsound.PlaySound("SystemExit", winsound.SND_ALIASES)
192
193
194 def get_urls(url: str) -> None:
195     global date_creation, flat_data
196     driver = get_driver()
197     try:
198         start_time = time.time()
199         full_url = url
200         driver.get(full_url)
201         driver._web_element_cls = uc.UCWebElement
202
203         try:
204             captcha = driver.find_elements(By.XPATH, xpath_dict["captcha"])
205             if captcha[-1] != []:
206                 print("Captcha")
207                 melody()
208                 time.sleep(25)
209         except:
210             pass
211
212         try:
213             from_date = driver.find_element(
214                 By.XPATH, xpath_dict["from_date"]

```

```

214         By.XPATH, xpath_dict["from_date"]
215     ).get_attribute("innerText")
216 except:
217     try:
218         from_date = driver.find_element(
219             By.XPATH, xpath_dict["from_date2"]
220         ).get_attribute("innerText")
221     except:
222         from_date = ""
223
224 try:
225     title = driver.find_element(By.XPATH, xpath_dict["title"]).get_attribute(
226         "innerText"
227     )
228 except:
229     try:
230         title = driver.find_element(
231             By.XPATH, xpath_dict["title2"]
232         ).get_attribute("innerText")
233     except:
234         title = ""
235
236 try:
237     JK = driver.find_element(By.XPATH, xpath_dict["JK"]).get_attribute(
238         "innerText"
239     )
240 except:
241     try:
242         JK = driver.find_element(By.XPATH, xpath_dict["JK2"]).get_attribute(
243             "innerText"
244         )
245     except:
246         JK = ""
247
248 try:
249     building = driver.find_element(
250         By.XPATH, xpath_dict["building"]
251     ).get_attribute("innerText")
252 except:
253     try:
254         building = driver.find_element(
255             By.XPATH, xpath_dict["building2"]
256         ).get_attribute("innerText")
257     except:
258         building = ""
259
260 try:
261     date_readiness = driver.find_element(
262         By.XPATH, xpath_dict["date_readiness"]
263     ).get_attribute("innerText")
264 except:
265     try:
266         date_readiness = driver.find_element(
267             By.XPATH, xpath_dict["date_readiness2"]
268         ).get_attribute("innerText")
269     except:
270         date_readiness = ""
271
272 try:
273     full_price = driver.find_element(
274         By.XPATH, xpath_dict["full_price"]
275     ).get_attribute("innerText")
276 except:
277     try:
278         full_price = driver.find_element(
279             By.XPATH, xpath_dict["full_price2"]
280         ).get_attribute("innerText")
281     except:
282         full_price = ""
283
284 try:
285     price_sq = driver.find_element(
286         By.XPATH, xpath_dict["price_sq"]
287     ).get_attribute("innerText")
288 except:
289     try:
290         price_sq = driver.find_element(
291             By.XPATH, xpath_dict["price_sq2"]
292         ).get_attribute("innerText")
293     except:
294         price_sq = ""
295
296 try:
297     metro_station = driver.find_element(
298         By.XPATH, xpath_dict["metro_station"]
299     ).get_attribute("innerText")
300 except:
301     try:
302         metro_station = driver.find_element(
303             By.XPATH, xpath_dict["metro_station2"]
304         ).get_attribute("innerText")
305     except:
306         try:
307             metro_station = driver.find_element(
308                 By.XPATH, xpath_dict["metro_station3"]
309             ).get_attribute("innerText")
310         except:
311             try:
312                 metro_station = driver.find_element(
313                     By.XPATH, xpath_dict["metro_station4"]
314                 ).get_attribute("innerText")
315             except:
316                 metro_station = ""
317
318 try:
319     metro_distance = driver.find_element(
320         By.XPATH, xpath_dict["metro_distance"]
321     ).get_attribute("innerText")

```

```

322:         except:
323:             try:
324:                 metro_distance = driver.find_element(
325:                     By.XPATH, xpath_dict["metro_distance2"]
326:                 ).get_attribute("innerText")
327:             except:
328:                 try:
329:                     metro_distance = driver.find_element(
330:                         By.XPATH, xpath_dict["metro_distance3"]
331:                     ).get_attribute("innerText")
332:                 except:
333:                     try:
334:                         metro_distance = driver.find_element(
335:                             By.XPATH, xpath_dict["metro_distance4"]
336:                         ).get_attribute("innerText")
337:                     except:
338:                         metro_distance = ""
339:
340:     try:
341:         city = driver.find_element(By.XPATH, xpath_dict["city"]).get_attribute(
342:             "innerText"
343:         )
344:     except:
345:         try:
346:             city = driver.find_element(By.XPATH, xpath_dict["city2"]).get_attribute(
347:                 "innerText"
348:             )
349:         except:
350:             city = ""
351:
352:     try:
353:         district = driver.find_element(
354:             By.XPATH, xpath_dict["district"]
355:         ).get_attribute("innerText")
356:     except:
357:         try:
358:             district = driver.find_element(
359:                 By.XPATH, xpath_dict["district2"]
360:             ).get_attribute("innerText")
361:         except:
362:             district = ""
363:
364:     try:
365:         area = driver.find_element(By.XPATH, xpath_dict["area"]).get_attribute(
366:             "innerText"
367:         )
368:     except:
369:         try:
370:             area = driver.find_element(By.XPATH, xpath_dict["area2"]).get_attribute(
371:                 "innerText"
372:             )
373:         except:
374:             area = ""
375:
376:     try:
377:         street = driver.find_element(By.XPATH, xpath_dict["street"]).get_attribute(
378:             "innerText"
379:         )
380:     except:
381:         try:
382:             street = driver.find_element(
383:                 By.XPATH, xpath_dict["street2"]
384:             ).get_attribute("innerText")
385:         except:
386:             street = ""
387:
388:     try:
389:         house = driver.find_element(By.XPATH, xpath_dict["house"]).get_attribute(
390:             "innerText"
391:         )
392:     except:
393:         try:
394:             house = driver.find_element(
395:                 By.XPATH, xpath_dict["house2"]
396:             ).get_attribute("innerText")
397:         except:
398:             house = ""
399:
400:     try:
401:         full_square = driver.find_element(
402:             By.XPATH, xpath_dict["full_square"]
403:         ).get_attribute("innerText")
404:     except:
405:         try:
406:             full_square = driver.find_element(
407:                 By.XPATH, xpath_dict["full_square2"]
408:             ).get_attribute("innerText")
409:         except:
410:             try:
411:                 full_square = driver.find_element(
412:                     By.XPATH, xpath_dict["full_square3"]
413:                 ).get_attribute("innerText")
414:             except:
415:                 full_square = ""
416:
417:     try:
418:         living_square = driver.find_element(
419:             By.XPATH, xpath_dict["living_square"]
420:         ).get_attribute("innerText")
421:     except:
422:         try:
423:             living_square = driver.find_element(
424:                 By.XPATH, xpath_dict["living_square2"]
425:             ).get_attribute("innerText")
426:         except:
427:             try:
428:                 living_square = driver.find_element(

```

```

429         By.XPATH, xpath_dict["living_square3"])
430     ).get_attribute("innerText")
431 except:
432     living_square = ""
433
434 try:
435     kitchen_square = driver.find_element(
436         By.XPATH, xpath_dict["kitchen_square"]
437     ).get_attribute("innerText")
438 except:
439     try:
440         kitchen_square = driver.find_element(
441             By.XPATH, xpath_dict["kitchen_square2"]
442         ).get_attribute("innerText")
443     except:
444         try:
445             kitchen_square = driver.find_element(
446                 By.XPATH, xpath_dict["kitchen_square3"]
447             ).get_attribute("innerText")
448         except:
449             kitchen_square = ""
450
451 try:
452     floor = driver.find_element(By.XPATH, xpath_dict["floor"]).get_attribute(
453         "innerText"
454     )
455 except:
456     try:
457         floor = driver.find_element(
458             By.XPATH, xpath_dict["floor2"]
459         ).get_attribute("innerText")
460     except:
461         try:
462             floor = driver.find_element(
463                 By.XPATH, xpath_dict["floor3"]
464             ).get_attribute("innerText")
465         except:
466             floor = ""
467
468 try:
469     finishing = driver.find_element(
470         By.XPATH, xpath_dict["finishing"]
471     ).get_attribute("innerText")
472 except:
473     try:
474         finishing = driver.find_element(
475             By.XPATH, xpath_dict["finishing2"]
476         ).get_attribute("innerText")
477     except:
478         try:
479             finishing = driver.find_element(
480                 By.XPATH, xpath_dict["finishing3"]
481             ).get_attribute("innerText")
482         except:
483             try:
484                 finishing = driver.find_element(
485                     By.XPATH, xpath_dict["finishing4"]
486                 ).get_attribute("innerText")
487             except:
488                 finishing = ""
489
490 try:
491     developer = driver.find_element(
492         By.XPATH, xpath_dict["developer"]
493     ).get_attribute("innerText")
494 except:
495     try:
496         developer = driver.find_element(
497             By.XPATH, xpath_dict["developer2"]
498         ).get_attribute("innerText")
499     except:
500         try:
501             developer = driver.find_element(
502                 By.XPATH, xpath_dict["developer3"]
503             ).get_attribute("innerText")
504         except:
505             developer = ""
506
507 try:
508     info_about_flat_1 = driver.find_element(
509         By.XPATH, xpath_dict["info about flat_1"]
510     ).get_attribute("innerText")
511 except:
512     try:
513         info_about_flat_1 = driver.find_element(
514             By.XPATH, xpath_dict["info about flat_11"]
515         ).get_attribute("innerText")
516     except:
517         try:
518             info_about_flat_1 = driver.find_element(
519                 By.XPATH, xpath_dict["info about flat_12"]
520             ).get_attribute("innerText")
521         except:
522             info_about_flat_1 = ""
523
524 try:
525     info_about_flat_2 = driver.find_element(
526         By.XPATH, xpath_dict["info about flat_2"]
527     ).get_attribute("innerText")
528 except:
529     try:
530         info_about_flat_2 = driver.find_element(
531             By.XPATH, xpath_dict["info about flat_21"]
532         ).get_attribute("innerText")
533     except:
534         try:
535             info_about_flat_2 = driver.find_element(

```

```
536         By.XPATH, xpath_dict["info about flat_22"]
537     ).get_attribute("innerText")
538 except:
539     info_about_flat_2 = ""
540
541 try:
542     info_about_flat_3 = driver.find_element(
543         By.XPATH, xpath_dict["info about flat_3"]
544     ).get_attribute("innerText")
545 except:
546     try:
547         info_about_flat_3 = driver.find_element(
548             By.XPATH, xpath_dict["info about flat_31"]
549         ).get_attribute("innerText")
550     except:
551         try:
552             info_about_flat_3 = driver.find_element(
553                 By.XPATH, xpath_dict["info about flat_32"]
554             ).get_attribute("innerText")
555         except:
556             info_about_flat_3 = ""
557
558 try:
559     info_about_flat_4 = driver.find_element(
560         By.XPATH, xpath_dict["info about flat_4"]
561     ).get_attribute("innerText")
562 except:
563     try:
564         info_about_flat_4 = driver.find_element(
565             By.XPATH, xpath_dict["info about flat_41"]
566         ).get_attribute("innerText")
567     except:
568         try:
569             info_about_flat_4 = driver.find_element(
570                 By.XPATH, xpath_dict["info about flat_42"]
571             ).get_attribute("innerText")
572         except:
573             info_about_flat_4 = ""
574
575 try:
576     info_about_flat_5 = driver.find_element(
577         By.XPATH, xpath_dict["info about flat_5"]
578     ).get_attribute("innerText")
579 except:
580     try:
581         info_about_flat_5 = driver.find_element(
582             By.XPATH, xpath_dict["info about flat_51"]
583         ).get_attribute("innerText")
584     except:
585         try:
586             info_about_flat_5 = driver.find_element(
587                 By.XPATH, xpath_dict["info about flat_52"]
588             ).get_attribute("innerText")
589         except:
590             info_about_flat_5 = ""
591
592 try:
593     info_about_flat_6 = driver.find_element(
594         By.XPATH, xpath_dict["info about flat_6"]
595     ).get_attribute("innerText")
596 except:
597     try:
598         info_about_flat_6 = driver.find_element(
599             By.XPATH, xpath_dict["info about flat_61"]
600         ).get_attribute("innerText")
601     except:
602         try:
603             info_about_flat_6 = driver.find_element(
604                 By.XPATH, xpath_dict["info about flat_62"]
605             ).get_attribute("innerText")
606         except:
607             info_about_flat_6 = ""
608
609 try:
610     info_about_flat_7 = driver.find_element(
611         By.XPATH, xpath_dict["info about flat_7"]
612     ).get_attribute("innerText")
613 except:
614     try:
615         info_about_flat_7 = driver.find_element(
616             By.XPATH, xpath_dict["info about flat_71"]
617         ).get_attribute("innerText")
618     except:
619         try:
620             info_about_flat_7 = driver.find_element(
621                 By.XPATH, xpath_dict["info about flat_72"]
622             ).get_attribute("innerText")
623         except:
624             info_about_flat_7 = ""
625
626 try:
627     info_about_flat_8 = driver.find_element(
628         By.XPATH, xpath_dict["info about flat_8"]
629     ).get_attribute("innerText")
630 except:
631     try:
632         info_about_flat_8 = driver.find_element(
633             By.XPATH, xpath_dict["info about flat_81"]
634         ).get_attribute("innerText")
635     except:
636         try:
637             info_about_flat_8 = driver.find_element(
638                 By.XPATH, xpath_dict["info about flat_82"]
639             ).get_attribute("innerText")
640         except:
641             info_about_flat_8 = ""
642
643 try:
```

```

643     try:
644         info_about_flat_9 = driver.find_element(
645             By.XPATH, xpath_dict["info about flat_9"]
646         ).get_attribute("innerText")
647     except:
648         try:
649             info_about_flat_9 = driver.find_element(
650                 By.XPATH, xpath_dict["info about flat_91"]
651             ).get_attribute("innerText")
652         except:
653             try:
654                 info_about_flat_9 = driver.find_element(
655                     By.XPATH, xpath_dict["info about flat_92"]
656                 ).get_attribute("innerText")
657             except:
658                 info_about_flat_9 = ""
659
660     try:
661         info_about_house_1 = driver.find_element(
662             By.XPATH, xpath_dict["info about house_1"]
663         ).get_attribute("innerText")
664     except:
665         try:
666             info_about_house_1 = driver.find_element(
667                 By.XPATH, xpath_dict["info about house_11"]
668             ).get_attribute("innerText")
669         except:
670             info_about_house_1 = ""
671
672     try:
673         info_about_house_2 = driver.find_element(
674             By.XPATH, xpath_dict["info about house_2"]
675         ).get_attribute("innerText")
676     except:
677         try:
678             info_about_house_2 = driver.find_element(
679                 By.XPATH, xpath_dict["info about house_21"]
680             ).get_attribute("innerText")
681         except:
682             info_about_house_2 = ""
683
684     try:
685         info_about_house_3 = driver.find_element(
686             By.XPATH, xpath_dict["info about house_3"]
687         ).get_attribute("innerText")
688     except:
689         try:
690             info_about_house_3 = driver.find_element(
691                 By.XPATH, xpath_dict["info about house_31"]
692             ).get_attribute("innerText")
693         except:
694             info_about_house_3 = ""
695
696     try:
697         elem_wait = WebDriverWait(driver, 5).until(
698             EC.presence_of_element_located(
699                 (By.XPATH, xpath_dict["start price button"])
700             )
701         )
702         action = ActionChains(driver, duration=500)
703         move = driver.find_element(By.XPATH, xpath_dict["start price button"])
704         action.move_to_element(move).perform()
705         elem_wait = WebDriverWait(driver, 6, 3).until(
706             EC.presence_of_element_located((By.XPATH, xpath_dict["start price"]))
707         )
708         start_price = driver.find_element(
709             By.XPATH, xpath_dict["start price"]
710         ).get_attribute("innerText")
711     except:
712         try:
713             elem_wait = WebDriverWait(driver, 5).until(
714                 EC.presence_of_element_located(
715                     (By.XPATH, xpath_dict["start price button2"])
716                 )
717             )
718             action = ActionChains(driver, duration=500)
719             move = driver.find_element(By.XPATH, xpath_dict["start price button2"])
720             action.move_to_element(move).perform()
721             elem_wait = WebDriverWait(driver, 6, 3).until(
722                 EC.presence_of_element_located(
723                     (By.XPATH, xpath_dict["start price2"])
724                 )
725             )
726             start_price = driver.find_element(
727                 By.XPATH, xpath_dict["start price2"]
728             ).get_attribute("innerText")
729         except:
730             start_price = ""
731
732     try:
733         start_date = driver.find_element(
734             By.XPATH, xpath_dict["start date"]
735         ).get_attribute("innerText")
736     except:
737         try:
738             start_date = driver.find_element(
739                 By.XPATH, xpath_dict["start date2"]
740             ).get_attribute("innerText")
741         except:
742             start_date = ""
743
744     try:
745         driver.find_elements(By.XPATH, xpath_dict["button_date_creation"])[
746             -1
747         ].click()
748         elem = WebDriverWait(driver, 9, 3).until(
749             EC.visibility_of_element_located(
750                 (By.XPATH, xpath_dict["date_creation"])

```

```

750         (By.XPATH, xpath_dict["date_creation"])
751     )
752 )
753 date_creation = driver.find_element(
754     By.XPATH, xpath_dict["date_creation"]
755 ).get_attribute("innerText")
756 except:
757     try:
758         driver.find_elements(By.XPATH, xpath_dict["button_date_creation2"])[
759             -1
760         ].click()
761         elem = WebDriverWait(driver, 9, 3).until(
762             EC.visibility_of_element_located(
763                 (By.XPATH, xpath_dict["date_creation"])
764             )
765         )
766         date_creation = driver.find_element(
767             By.XPATH, xpath_dict["date_creation"]
768         ).get_attribute("innerText")
769     except:
770         date_creation = ""
771
772 flat_data = (
773     full_url,
774     from_date,
775     title,
776     JK,
777     start_price,
778     start_date,
779     building,
780     date_readiness,
781     full_price,
782     price_sq,
783     metro_station,
784     metro_distance,
785     city,
786     district,
787     area,
788     street,
789     house,
790     full_square,
791     living_square,
792     kitchen_square,
793     floor,
794     finishing,
795     developer,
796     date_creation,
797     info_about_flat_1,
798     info_about_flat_2,
799     info_about_flat_3,
800     info_about_flat_4,
801     info_about_flat_5,
802     info_about_flat_6,
803     info_about_flat_7,
804     info_about_flat_8,
805     info_about_flat_9,
806     info_about_house_1,
807     info_about_house_2,
808     info_about_house_3,
809 )
810
811 try:
812     unpublished = driver.find_element(By.XPATH, xpath_dict["unpublished"])
813     if unpublished.get_attribute("innerText").__contains__(" "):
814         flat_data = tuple(
815             [full_url, "unpublished"] + [""] * (len(flat_data) - 2)
816         )
817 except:
818     try:
819         unpublished = driver.find_element(By.XPATH, xpath_dict["unpublished2"])
820         if unpublished.get_attribute("innerText").__contains__(" "):
821             flat_data = tuple(
822                 [full_url, "unpublished"] + [""] * (len(flat_data) - 2)
823             )
824     except:
825         pass
826
827 with open(rf"{my_path}\raw_data.csv", "a", encoding="utf-8", newline="") as ouf:
828     writer = csv.writer(
829         ouf,
830         delimiter="$",
831         quotechar="|",
832         quoting=csv.QUOTE_MINIMAL,
833     )
834     writer.writerow(flat_data)
835
836 print(
837     f"{full_url[-15:]}, {(time.time()-start_time)//60:.0f}min {(time.time()-start_time)%60:.2f}s; "
838     f"in total {(time.time()-whole_time)//3600:.0f}h {(time.time()-whole_time)%3600//60:.0f}min {(time.time()-whole_time)%60:.2f}s"
839 )
840
841 except Exception as ex:
842     print(ex)
843     print(f"Error with -> {full_url}")
844
845
846 if __name__ == "__main__":
847     p = Pool(processes=2)
848     p.map(get_urls, url_list)

```