

Introduction

This is a scrapy data extraction assignment where you need to use [Scrapy Framework](#) for data extraction.

The objective of this task is to evaluate the learning, technical and other skills related to a programming environment.

For the given website the candidate must do the following guidelines to extract the data and store in the required format mentioned below.

1. The candidate must develop a scrapy project for the website provided and extract a minimum of 1000 data items from the website and store the data in a database
2. The project should be well structured and modularized.
3. The coding must go through three important steps:
 - a. Crawling
 - i. Going through each of the URLs from the URL provided and going through each and every pages as well (Proper Pagination)
 - b. Parsing
 - i. Parsing is to be done in the last depth, where we find product/person/property details, all of the required fields mentioned below is to be collected using xpath
 - c. Cleaning & Data Structuring
 - i. The extracted data should be cleaned properly and the data should be structured in the specified format as explained below.

Prerequisite:

1. **Scrapy framework should be used.**
2. **The data extracted should be in JSON file format and there must be at least 1000 data items extracted from the website.**
3. **Should submit the code via private git repo and dropbox URL of the output json file.**
4. **The code should be in Python3.**
5. **Code should be optimized and reliable.**
6. **Must follow [PEP8](#) standards for the code.**
7. **Read the below instructions carefully and complete the spider based on the exact requirements.**

Task

Website to scrape : <https://www.bayut.com/>



Total time required to complete: 10 days


Start URL: <https://www.bayut.com/to-rent/property/dubai/> [The data extraction process should start from this URL]

Refer following images for guidelines

A. Crawling

1. The crawler should go through each and every property URLs as marked in the image, as in the URL above if there are 24 URLs in the page, it should visit all of them.

Furnish Type: ☒ All ☐ Furnished ☐ Unfurnished ☐ TruCheck™ first  









TruCheck™

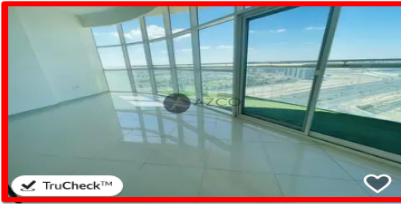
AED 60,000 Yearly

Noora Residence 1, Noora Residence, Jumeirah Villa...
Apartment

Cheapest Deal | Chiller Free | Massiv...

 2  3  2,744 sqft









TruCheck™


AED 45,000 Yearly

Reef Residence, Jumeirah Village Circle (JVC), Dubai
Apartment

Fully Panoramic| Glass Building| Bur...

 1  2  881 sqft









TruCheck™

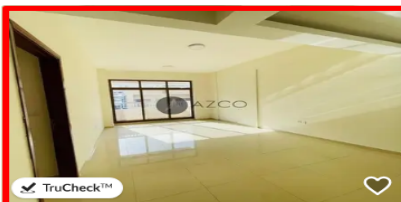
AED 64,000 Yearly

Tasmeer Residence, Jumeirah Village Circle (JVC), D...
Apartment

Chiller Free 2 Bedroom! Pool View! ...

 2  2  1,234 sqft









TruCheck™

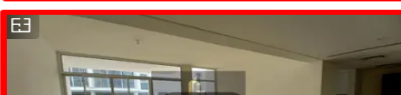
AED 45,000 Yearly

Laya Residences, Jumeirah Village Circle (JVC), Dubai
Apartment

Spacious| Maids Room| Near To Scho...

 1  2  1,000 sqft



TruCheck™

AED 110,000 Yearly

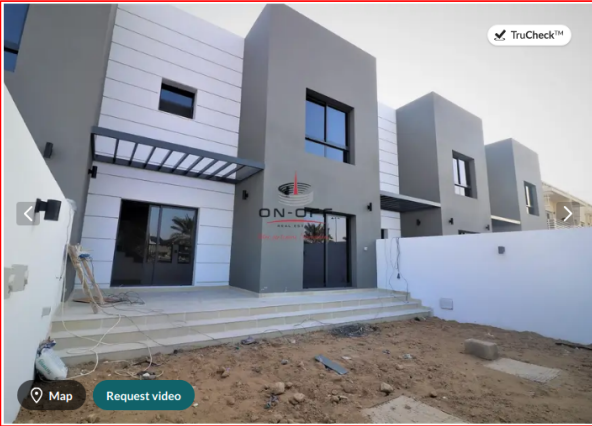
Arabella 1, Arabella Townhouses, Mudon, Dubai
Townhouse

2. The crawler should go through each and every page of the property listing page. i.e. if there are 1500 pages, it should visit all of 1500 pages and every property URLs in each page.



B. Parsing

< Back to Search Bayut 12 Townhouses for Rent in Dubai > Al Furjan > Phase 2 Townhouses > Bayut - 2335-Th-R-0039



11

On off Real Estate L.L.C
DED# 876441
RERA# 23478 9
Permit# 2511742192
Agent: Jahangir Ali 10

View all properties >

Contact Agent for more information.

NAME*

EMAIL*

PHONE*

+971

MESSAGE*

I would like to inquire about your property Bayut - 2335-Th-R-0039. Please contact me at your earliest convenience.

☒ Keep me informed about similar properties.

CALL SEND EMAIL

Phase 2
See the community attractions and lifestyle

Useful Links

Properties for rent in Dubai
Properties for rent in Al Furjan
Properties for rent in Phase 2
Townhouses for rent in Dubai
Townhouses for rent in Al Furjan
Townhouses for rent in Phase 2
4 Bedroom Townhouses for rent in Dubai
4 Bedroom Townhouses for rent in Al Furjan
4 Bedroom Townhouses for rent in Phase 2

Report this property

AED 99,999⁶ Yearly

Phase 2, Al Furjan, Dubai⁷

4 Beds 6 Baths 3,100 sqft⁸

Overview Location & Nearby Trends

All Ensuite Bedroom|Near Spinneys|Phase2¹⁴

On Off Properties proud to offer you this amazing Huge size 4 bedrooms villa in phase 2 Al Furjan.

All Bedrooms with Attached Baths
Very near to Furjan Pavilion Phase 2 Villas
Description

Read More

Type Townhouse³ Furnishing Unfurnished⁵

Purpose For Rent² TruCheck™ on 29 April 2021

Reference no. Bayut - 2335-Th-R-0039¹ Added on 22 April 2021⁴

Features / Amenities

Balcony or Terrace	Maids Room	Centrally Air-Conditioned	Storage Areas	Study Room
Kids Play Area	Lawn or Garden	Cafeteria or Canteen	Conference Room	+ 5 more amenities

13

The above image can be downloaded [here](#).

Expected single data output:

Eg URL: <https://www.bayut.com/property/details-5118159.html>

Check the above image for mapping fields to be extracted.

Sl. No	Field Name	Field Type	Example
1	property_id	string	"Bayut - 2335-Th-R-0039"
2	purpose	string	"For Rent"
3	type	string	"Townhouse"
4	added_on	string	"22 April 2021"
5	furnishing	string	"Unfurnished"
6	price	dict	{"currency": "AED", "amount": "99,999"}
7	location	string	"Phase 2, Al Furjan, Dubai"
8	bed_bath_size	dict	{"bedrooms": 4, "bathrooms": 6, "size": "3,100 sqft"}
9	permit_number	string	"6511742172"
10	agent_name	string	"Jahangir Ali"
11	image_url	string	"https://images.bayut.com/thumbnails/104428356-400x300.jpeg"
12	breadcrumbs	string	"Townhouses for Rent in Dubai > Al Furjan > Phase 2 Townhouses"
13	amenities	list	["Balcony or Terrace", "Maids Room", "Centrally Air-Conditioned", "Storage Areas", "Study Room", "Kids Play Area", "Lawn or Garden", "Cafeteria or Canteen", "Conference Room"]

14	description	string	"On Off Properties proud to offer you this amazing Huge size 4 bedrooms villa in phase 2 Al Furjan. All Bedrooms with Attached Baths Very near to Furjan Pavilion Phase 2 Villas Description All Bedroom with Attached Bath & Terrace One Bedroom on ground floor Fully Closed Kitchen Maids Room with Ensuite bathroom Laundry room Backyard Garden BUA 3100 sqft About the Location: - 10 mins to Ibne Battuta Mall - Easy access road to Sheikh Mohd Bin Zayed & Sheikh Zayed Road - Nearby Petrol Station And Super Markets Al Furjan is a vibrant residential development located between Sheikh Zayed Road and Mohammed Bin Zayed Road, adjacent to the Discovery Gardens community, conveniently close to Ibn Battuta Mall & with easy access back towards the Dubai Marina and other nearby prominent developments and is accessible from Al Yalayis Street and Al Asayel Street. We can also do online viewing by webcam in case youre overseas or youre following Stay home during this special period. For Viewing & Info Syed Jahangir"
----	-------------	--------	---

Example Structure of above data item:

```
{
  "property_id": "Bayut - 2335-Th-R-0039",
  "purpose": "For rent",
  "type": "Townhouse",
  "added_on": "22 April 2021",
  "furnishing": "Unfurnished",
  "price": {
    "currency": "AED",
    "amount": "99,999"
  },
  "location": "Phase 2, Al Furjan, Dubai",
  "bed_bath_size": {
    "bedrooms": 4,
    "bathrooms": 6,
    "size": "3,100 sqft"
  },
  "permit_number": "6511742172",
  "agent_name": "Jahangir Ali",
  "image_url":
  "https://images.bayut.com/thumbnails/104428356-400x300.jpeg",
  "breadcrumbs": "Townhouses for Rent in Dubai > Al Furjan > Phase 2
Townhouses",
  "amenities" : ["Balcony or Terrace", "Maids Room", "Centrally
Air-Conditioned", "Storage Areas", "Study Room", "Kids Play Area", "Lawn or
```

Garden", "Cafeteria or Canteen", "Conference Room"],

"description": "On Off Properties proud to offer you this amazing Huge size 4 bedrooms villa in phase 2 Al Furjan. All Bedrooms with Attached Baths Very near to Furjan Pavilion Phase 2 Villas Description All Bedroom with Attached Bath & Terrace One Bedroom on ground floor Fully Closed Kitchen Maids Room with Ensuite bathroom Laundry room Backyard Garden BUA 3100 sqft About the Location: - 10 mins to Ibne Battuta Mall - Easy access road to Sheikh Mohd Bin Zayed & Sheikh Zayed Road - Nearby Petrol Station And Super Markets Al Furjan is a vibrant residential development located between Sheikh Zayed Road and Mohammed Bin Zayed Road, adjacent to the Discovery Gardens community, conveniently close to Ibn Battuta Mall & with easy access back towards the Dubai Marina and other nearby prominent developments and is accessible from Al Yalaisy Street and Al Asayel Street. We can also do online viewing by webcam in case youre overseas or youre following Stay home during this special period. For Viewing & Info Syed Jahangir"

}