



Collecting & Analysing Data for Elissa



Bruxelles, 27/09/2020

Philippe-Fimmers, Francesco Mariottini and Opap's Ditudidi



From Objectives to Modules

OBJECTIVES

- 1) Collecting data for 10000 buildings.
- 2) Removing duplicates.
- 3) Replacing empty information with None.
- 4) Replacing True/False with “Yes or No.

INDIPENDENT MODULES

- a) `get_search_results` (1)
- b) `scrap_search_result` (1)
- c) `dataquality` (2, 3, 4)

Workflow
with CSV
creation









Scrapping results solutions

Mansion for sale
4 bedrooms | 330 m²
[Boulevard de la Dodaine 7 | 1400 — Nivelles](#)


€600,000
[Request your mortgage loan](#)
Immoweb code : 8930262





[photos](#) [map](#)


[Log in to see statistics about this property](#) [Report a mistake](#) [Print](#)

Overview

 4 bedrooms

 2 bathrooms

 330 m² livable space

 390 m² of land

Description

-

-

General

Available as of	Upon exchange of deeds
Construction year	1924
Building condition	Good
Street facade width	6 m
Facades	3
Covered parking spaces	1
Outdoor parking spaces	2
Virtual visit	Link to the virtual visit

House

Property sub-types | [Reset](#)

☐ Bungalow

☐ Castle

☐ Country cottage

☐ Apartment block

☐ Town-house

☐ Villa

☐ Country house

☐ Chalet

☐ Farmhouse

☐ Exceptional property

☐ Mixed-use building

☐ Mansion

☐ Other properties

☐ Pavilion

Apartment

Property sub-types | [Reset](#)

☐ Ground floor

☐ Triplex

☐ Penthouse

☐ Kot

☐ Duplex

☐ Studio

☐ Loft

☐ Service flat



Scrapping results solutions




House for sale

1 2 ... 333 >

Sort : Relevance ^

NEW BUILD



House
€990,000
6 bdr. · 485 m²
1150 Woluwe-Saint-Pierre

Entre la Maison Communale et La Rosette - 2 Logements

Relevance
Cheapest
Most expensive
Postal code
Newest

orimo
REAL ESTATE



{property_type}

{page_number}



Scrapping results solutions



https://www.immoweb.be/en/classified/new-real-estate-project-apartments/for-sale/huy/4500/8876897?searchId=5f6f5228e4151


IMMOWEB For sale For rent Valuate Agencies Services Place a

< Back to the results < Previous Next >

RESIDENCE GODIN HUY €175,000 - €275,000

1 - 3 bedrooms | 60 - 108 m²
Avenue Godin-Parmajan 1.14500 - Huy Immoweb code : 8876897

NEW REAL ESTATE PROJECT 36% SOLD



photos map

Log in to see statistics about this property Report a mistake Print

Overview

1 - 3 bedrooms 60 - 108 m² Available date : 12/31/2022

Description

Nouvelle résidence centre-ville de Huy

Sorry, there is no English translation. Magnifique résidence de haut standing située dans le centre-ville comprenant 14 appartements 1, 2 et 3 chambres avec terrasse, cave et car-port ou parking sous-sol et 3 surfaces de bureaux au rez-de-...

[Read more](#)

All properties

14 apartments (9 available)
€ 175,000 - € 275,000

€ 175,000	1 bedroom	60 sqm	Floor 1	>
€ 220,000	2 bedrooms	93 sqm	Floor 1	>
SOLD	2 bedrooms	93 sqm	Floor 2	



https://www.immoweb.be/en/classified/new-real-estate-project-ap



Scrapping results solutions



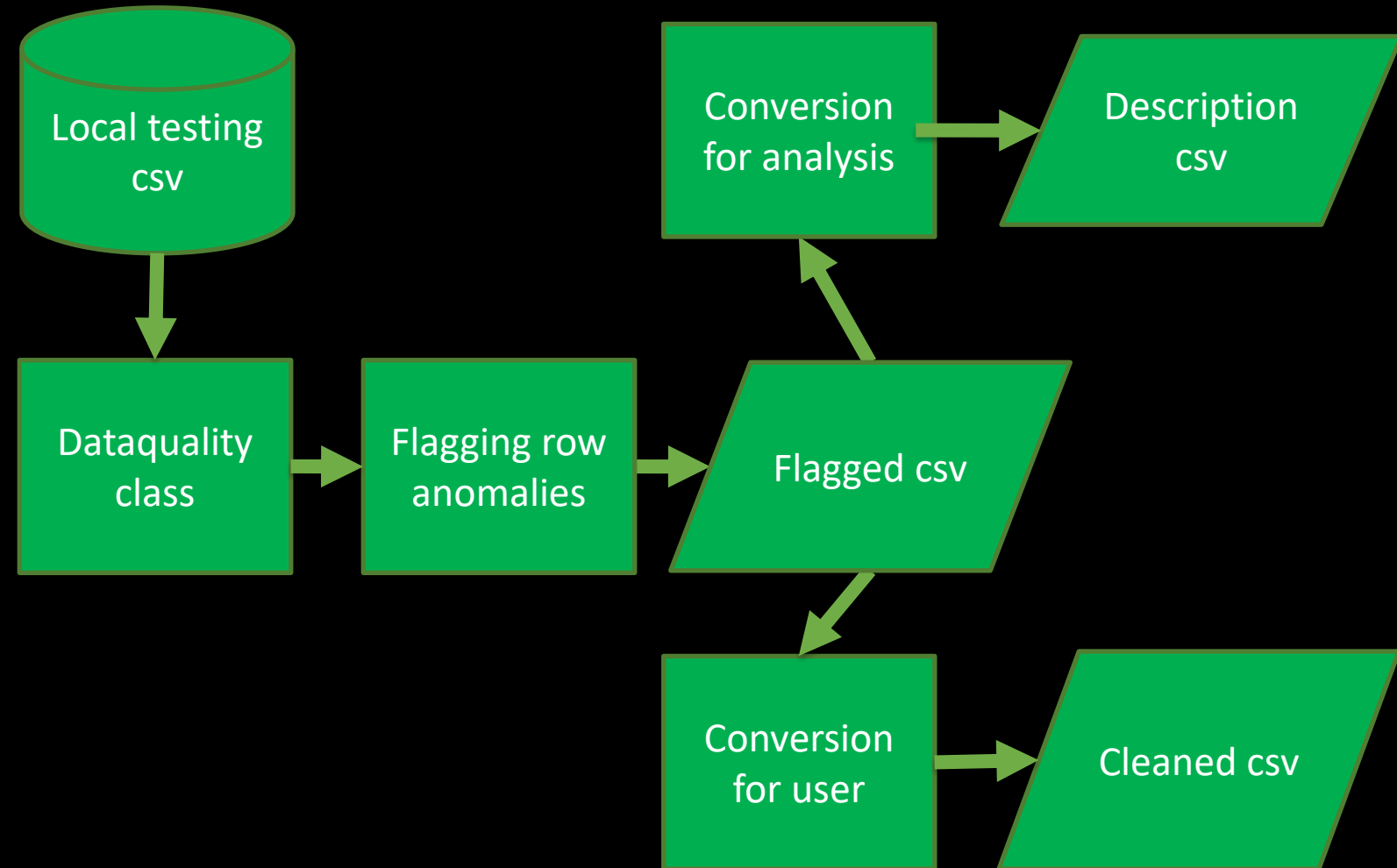
- Expected input : Dictionary {url string, is_house bool}
- Expected output : Dictionary {property string, value list}
- scrap_list : dispatches every url from the input list and builds the output dict
- scrap : receives one url and fetches the needed info



Dataquality solutions

CHALLENGES

- Building information not initially available.
- Building information not properly formatted.
- No explicit criteria on duplicates.
- Desired output (str) less effective for analysis.





Lesson learnt

- Clear input & output specifications.
- Internal workflow & deadlines.
- Working packages for parallel testing.
- Previous training for effective coding & support.

