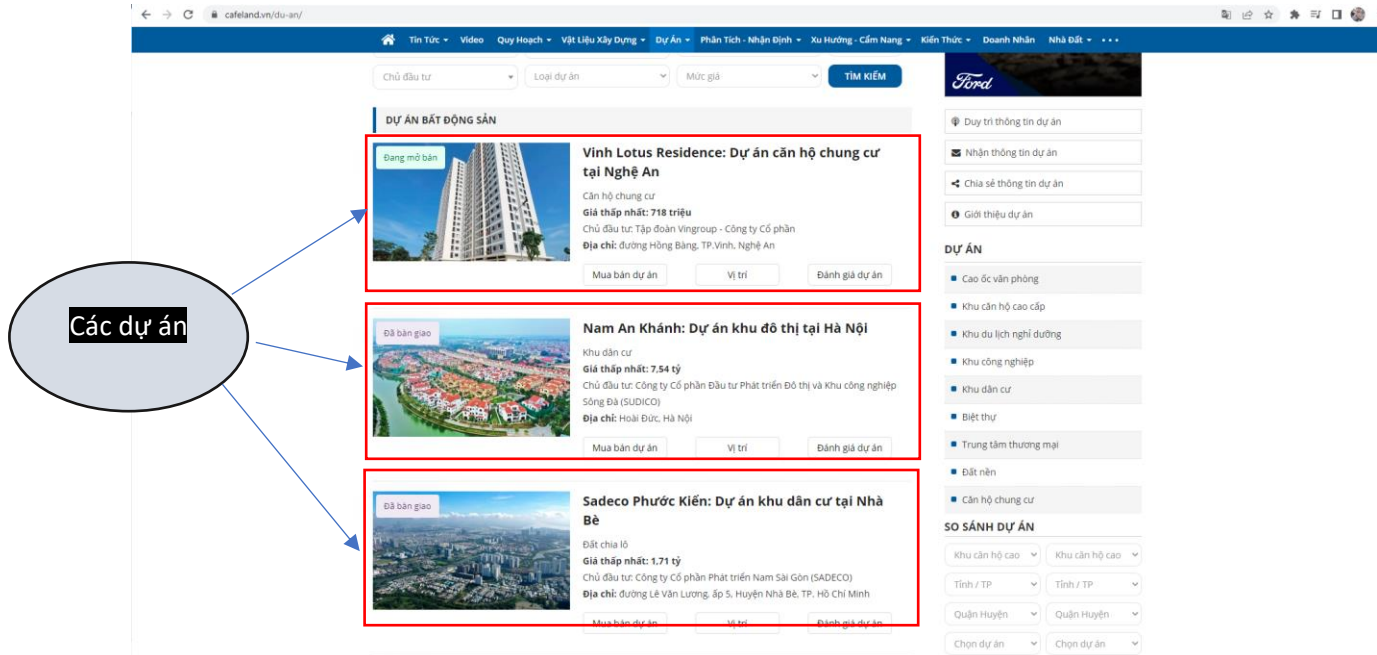


## BÀI TẬP CÀO DỮ LIỆU

**Bài tập 1** Yêu cầu sinh viên dùng thư viện Scrapy(python) để tiến hành cào dữ liệu bao gồm các yêu cầu sau :

1. Truy cập vào url : <https://cafeland.vn/du-an/>
  2. Thực hiện các tác vụ Scrapy lấy link các bài viết về dự án bất động sản
- \*\* Gợi ý :** Link nằm ở phần tên các dự án ( lấy attribute href )



Hình 1. Hình minh họa khi truy cập vào url ở trên

3. Sau khi lấy cái link bài viết tiến hành tiếp tục gửi requests để lấy thông tin . Các thông tin bao gồm :

- Loại dự án
- Tên dự án
- Vị trí tọa độ của công trình trên map
- Thông tin nhanh của dự án

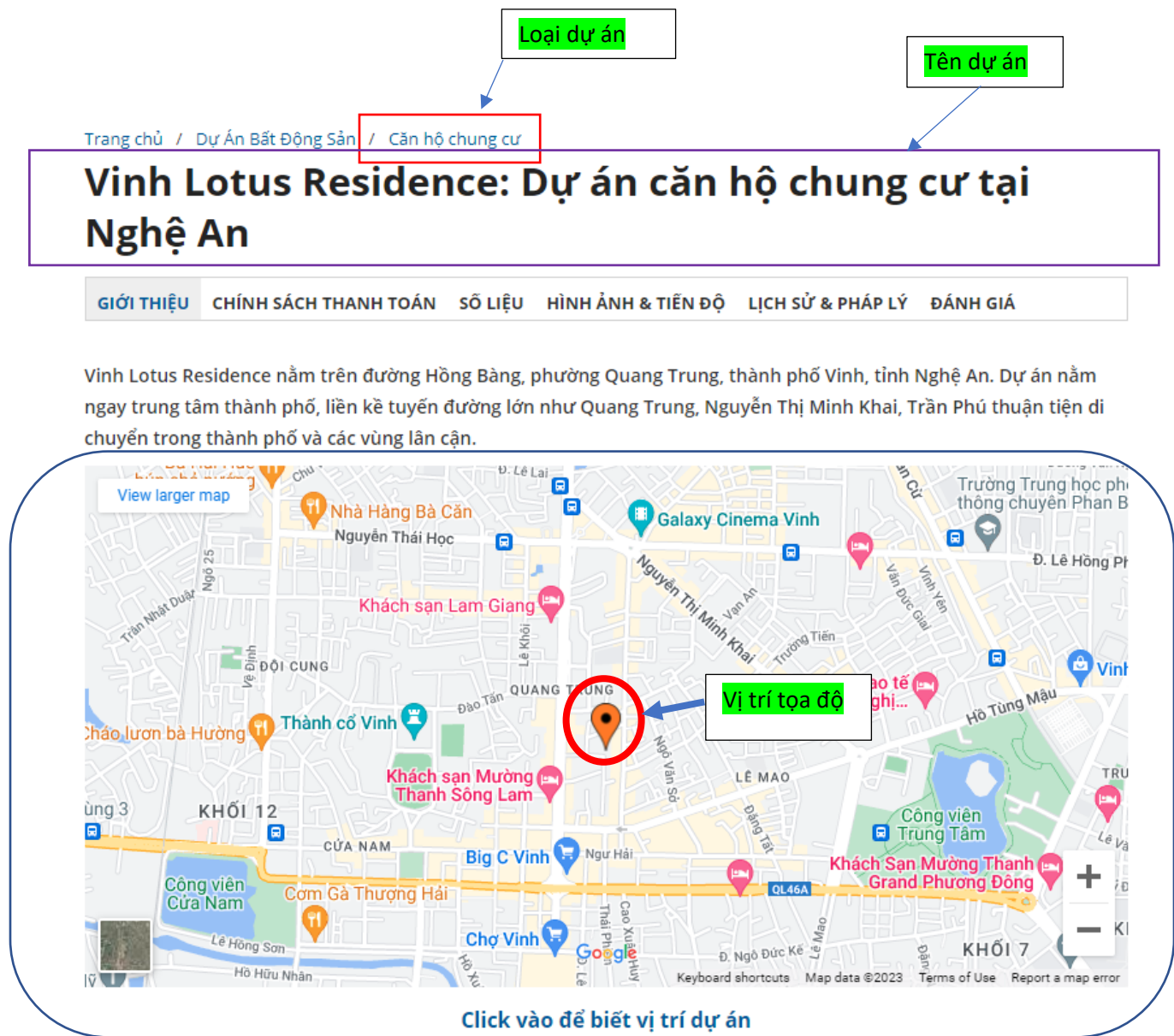
Đối với phần thông tin nhanh của dự án , sinh viên chỉ cần lấy 2 thông tin bao gồm **CHỦ ĐẦU TƯ** và **DIỆN TÍCH** ( ở hình 3 phía dưới có minh họa) . Nếu trong trường hợp thông tin đó không có hoặc bị thiếu thì sinh viên thay bằng chữ “ Không có ”

**\*\* Gợi ý :**

Loại dự án nằm trong thẻ **span** có attribute **itemprop** là **name**

Tên dự án nằm trong thẻ **h1** có attribute **class** là **“sevenPostTitle blue”**

Vị trí tọa độ map nằm trong **attribute href của thẻ a** ( thẻ a nào thì sinh viên tự tìm )



Hình 2 . Hình minh họa khi truy cập vào link của từng dự án .

\*\* Kéo chuột xuống phía dưới sẽ thấy thông tin nhanh của dự án

26 triệu đồng/m<sup>2</sup>.

### Thông tin nhanh dự án Vinhomes Lotus Residence Nghệ An

Dự án: Vinhomes Lotus Residence	Tổng diện tích: 900,93 m <sup>2</sup>
Vị trí: Đường Hồng Bàng, phường Quang Trung, thành phố Vinh, tỉnh Nghệ An	Tổng vốn đầu tư:
Loại hình: Căn hộ	Năm khởi công: Tháng 06/2019
Chủ đầu tư: Tập đoàn Vingroup – Công ty CP	Năm hoàn thành: Năm 2021

Hình 3 .Thông tin nhanh của dự án ( phần viền đỏ )

4. Thực hiện lấy link các bài dự án trong 10 page đầu ( next\_page) và làm các yêu cầu 1, 2, 3 ở trên
5. Kết quả trả ra file csv bao gồm các cột như bên dưới ( Sinh viên để ý tên các cột còn thông tin chỉ mang tính chất minh họa )

Loại dự án	Tên dự án	Vị trí tọa độ	Chủ đầu tư	Diện tích
Căn hộ chung cư	Vinh Lotus	18.670755377182- 105.67574589287	Vingroup	10 m2

Bảng 1 . Minh họa các cột dữ liệu sau khi cào cần trả ra

#### **YÊU CẦU :**

- Code sạch sẽ
- Xpath , Selector ngắn gọn ( nếu có thể )
- Có chú thích code (để người khác đọc)

**\*\* SINH VIÊN LÀM ĐẦY ĐỦ CÁC YÊU CẦU CỦA BÀI + TIÊU CHÍ CODE ĐƯỢC 1.5 ĐIỂM CÔNG**