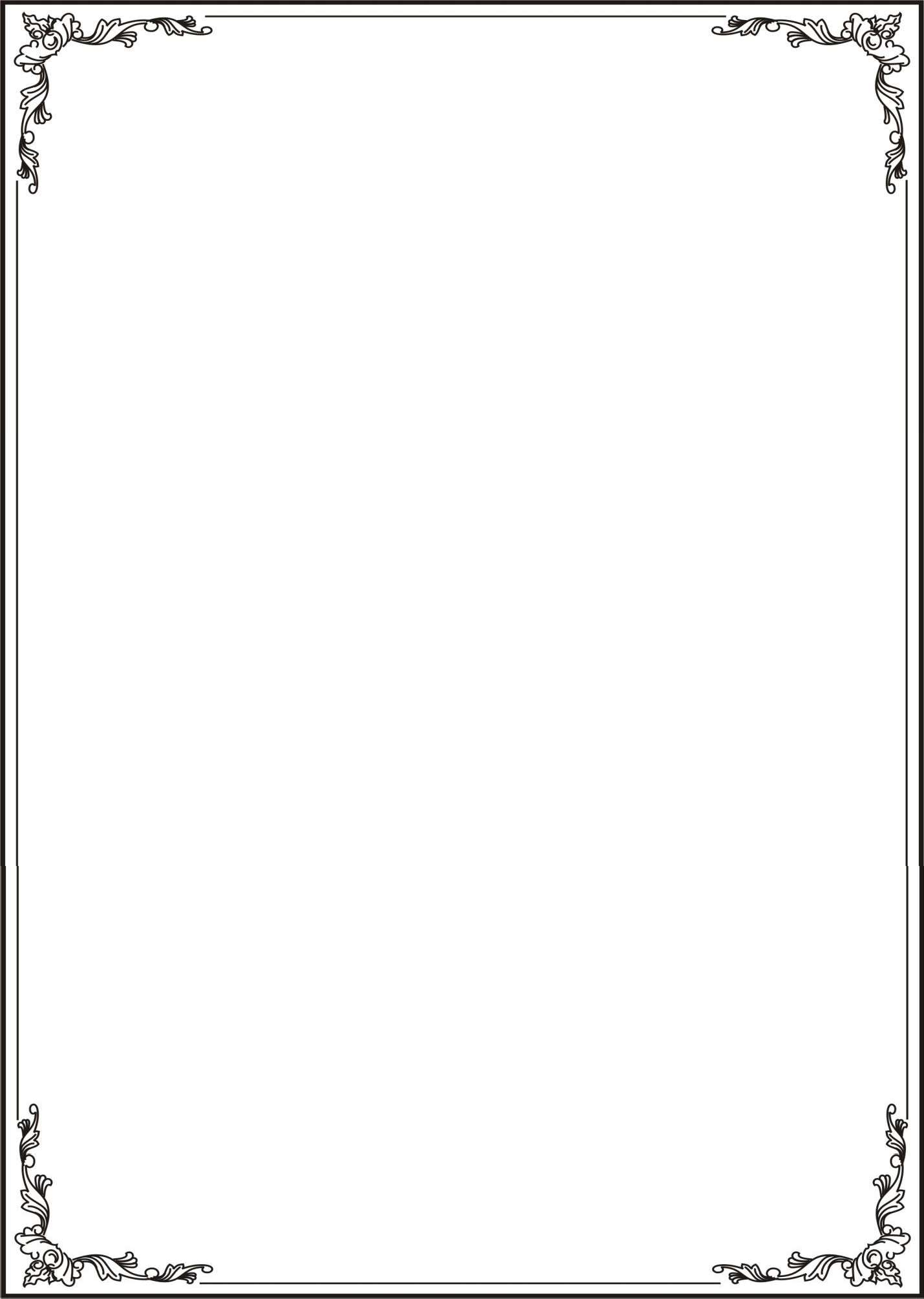
**BỘ GIÁO DỤC VÀ ĐÀO TẠO**



**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

**VŨ TRUNG KIÊN: 20133060**

**NGUYỄN VÕ SONG TOÀN: 20133097**

DỰ ĐOÁN GIÁ NHÀ VÀ CHUNG CƯ THÀNH PHỐ HỒ CHÍ MINH VỚI DỮ LIỆU LẤY TỪ NHATOT

**TIỂU LUẬN CHUYÊN NGÀNH**

**NGÀNH KỸ THUẬT DỮ LIỆU**

GIÁO VIÊN HƯỚNG DẪN

**PGS.TS Hoàng Văn Dũng**

**KHÓA 2020 – 2024**

**ĐH SƯ PHẠM KỸ THUẬT TP.HCM CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM**

**KHOA CNTT Độc lập – Tự do – Hạnh Phúc**

**PHIẾU NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

Họ và tên Sinh viên 1: Vũ Trung Kiên MSSV 1: 20133060

Họ và tên Sinh viên 2: Nguyễn Võ Song Toàn MSSV 2: 20133097

Ngành: Kỹ thuật dữ liệu.

Tên đề tài: Dự đoán giá nhà và chung cư thành phố Hồ Chí Minh với dữ liệu từ nhatot.

Họ và tên giáo viên hướng dẫn: PGS.TS Hoàng Văn Dũng

**NHẬN XÉT**

1. Về nội dung đề tài khối lượng thực hiện:

...................................................................................................................................... ......................................................................................................................................

2. Ưu điểm và khuyết điểm:

.....................................................................................................................................

.....................................................................................................................................

3. Đánh giá loại:

4. Điểm

Tp*. Hồ Chí Minh, ngày tháng năm 2023*

Giáo viên hướng dẫn

*(Ký & ghi rõ họ tên)*

**PHIẾU NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN**

Họ và tên Sinh viên 1: Vũ Trung Kiên MSSV 1: 20133060

Họ và tên Sinh viên 2: Nguyễn Võ Song Toàn MSSV 2: 20133097

Ngành: Kỹ thuật dữ liệu.

Tên đề tài: Dự đoán giá nhà và chung cư thành phố Hồ Chí Minh với dữ liệu từ nhatot.

Họ và tên giáo viên hướng dẫn: PGS.TS Hoàng Văn Dũng

**NHẬN XÉT**

1. Về nội dung đề tài khối lượng thực hiện:

..............................................................................................................................................................................................................................................................

2. Ưu điểm và khuyết điểm:

...............................................................................................................................

...............................................................................................................................

3. Đánh giá loại:

4. Điểm:

Tp*. Hồ Chí Minh, ngày tháng năm 2023*

Giáo viên phản biện

*(Ký & ghi rõ họ tên)*

**LỜI CAM ĐOAN**

Chúng em xin cam đoan rằng, khoá luận "Dự đoán giá nhà và chung cư thành phố Hồ Chí Minh với dữ liệu từ nhatot." Là công trình nghiên cứu của chúng em dưới sự hướng dẫn của PGS.TS Hoàng Văn Dũng, xuất phát từ nhu cầu thực tiễn và nguyện vọng tìm hiểu của bản thân chúng em.

Ngoại trừ kết quả tham khảo từ các công trình khác đã ghi rõ trong luận văn, các nội dung trình bày trong khoá luận này là kết quả nghiên cứu do chính chúng em thực hiện và kết quả của khoá luận chưa từng công bố trước đây dưới bất kỳ hình thức nào.

Hồ Chí Minh, ngày ... tháng ... năm 2023

Tác giả

**Vũ Trung Kiên Nguyễn Võ Song Toàn**

**Trường Đại học Sư Phạm Kỹ Thuật TP.HCM**

**Khoa: Công nghệ thông tin**

**ĐỀ CƯƠNG TIỂU LUẬN CHUYÊN NGÀNH**

Họ và tên SV thực hiện: Vũ Trung Kiên MSSV: 20133060

Họ và tên SV thực hiện: Nguyễn Võ Song Toàn MSSV: 20133097

Thời gian làm tiểu luận: Đến:

Chuyên ngành: Kỹ Thuật Dữ Liệu

Tên đề tài: Dự đoán giá nhà và chung cư thành phố Hồ Chí Minh với dữ liệu từ nhatot.

Họ và tên Giáo viên hướng dẫn: PGS.TS Hoàng Văn Dũng

KẾ HOẠCH THỰC HIỆN

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |

|  |  |
| --- | --- |
| Ý kiến giáo viên hướng dẫn (Ký & ghi rõ họ tên) | Ngày … tháng … năm 2023 Người viết  Vũ Trung Kiên |
|  | Nguyễn Võ Song Toàn |
|  |  |

# DANH MỤC HÌNH ẢNH

1. **Phần MỞ ĐẦU**
   1. Tính cấp thiết của đề tài
   2. Mục đích của đề tài
   3. Cách tiếp cận và phương pháp nghiên cứu

* Đối tượng nghiên cứu
* Phạm vi nghiên cứu

1. Phân tích những công trình có liên quan
2. Kết quả dự kiến đạt được
3. **Phần NỘI DUNG *(chia thành từng chương)***

***(chỉ trình bày các ý chính trong từng chương)***

1. Chương 1
   1. …
   2. …
2. Chương 2
   1. …
   2. …
3. Chương 3
   1. …
   2. …
4. Chương 4
   1. ..
   2. …
5. **Phần KẾT LUẬN**
6. Tài liệu tham khảo

[1]…

[2]…

Contents

[DANH MỤC HÌNH ẢNH 7](#_Toc149471843)

[PHẦN MỞ ĐẦU 11](#_Toc149471844)

[1. TÍNH CẤP THIẾT CỦA ĐỀ TÀI 11](#_Toc149471845)

[2. MỤC ĐÍCH CỦA ĐỀ TÀI 11](#_Toc149471846)

[3. CÁCH TIẾP CẬN VÀ PHƯƠNG PHÁP NGHIÊN CỨU 12](#_Toc149471847)

[3.1. Đối tượng nghiện cứu 12](#_Toc149471848)

[3.2. Phạm vi nghiên cứu 12](#_Toc149471849)

[3.3. Cách tiếp cận và phương pháp nghiên cứu 12](#_Toc149471850)

[4. KẾT QUẢ VÀ DỰ KIẾN ĐẠT ĐƯỢC 13](#_Toc149471851)

[5. BỐ CỤC BÀI BÁO CÁO 13](#_Toc149471852)

[PHẦN NỘI DUNG 14](#_Toc149471853)

[CHƯƠNG 1: TỔNG QUAN VỀ THỊ TRƯỜNG BẤT ĐỘNG SẢN TẠI THÀNH PHỐ HỒ CHÍ MINH 14](#_Toc149471854)

[1.1. Giới thiệu 14](#_Toc149471855)

[1.2 Phát triển kinh tế và tăng trưởng dân số 14](#_Toc149471856)

[1.3. Tình hình cung cầu bất động sản 14](#_Toc149471857)

[1.4. Các yếu tố ảnh hưởng đến giá nhà và chung cư 15](#_Toc149471858)

[CHƯƠNG 2. CÁC PHƯƠNG PHÁP DỰ ĐOÁN GIÁ NHÀ VÀ CHUNG CƯ 15](#_Toc149471859)

[2.1. Phân tích hồi quy tuyến tính 15](#_Toc149471860)

[B. Mô hình dự báo thống kê 19](#_Toc149471861)

[C. Mạng nơ-ron nhân tạo 20](#_Toc149471862)

[D. Các phương pháp khác 20](#_Toc149471863)

[IV. Thu thập dữ liệu và xử lý dữ liệu 20](#_Toc149471864)

[A. Nguồn dữ liệu 20](#_Toc149471865)

[B. Phương pháp thu thập dữ liệu 20](#_Toc149471866)

[C. Tiền xử lý dữ liệu 20](#_Toc149471867)

[V. Xây dựng mô hình dự đoán giá nhà và chung cư 20](#_Toc149471868)

[A. Mô tả mô hình 20](#_Toc149471869)

[B. Huấn luyện mô hình 20](#_Toc149471870)

[C. Đánh giá và tinh chỉnh mô hình 20](#_Toc149471871)

[VI. Dự đoán giá nhà và chung cư tại thành phố Hồ Chí Minh 20](#_Toc149471872)

[A. Kết quả dự đoán 20](#_Toc149471873)

[B. So sánh với giá thực tế 20](#_Toc149471874)

[C. Đánh giá hiệu suất và độ chính xác của mô hình 20](#_Toc149471875)

[VII. Thảo luận và kết luận 20](#_Toc149471876)

[A. Thảo luận kết quả 20](#_Toc149471877)

[B. Đóng góp của nghiên cứu 20](#_Toc149471878)

[C. Hạn chế và hướng phát triển tiếp theo 20](#_Toc149471879)

[VIII. Tài liệu tham khảo 20](#_Toc149471880)

# PHẦN MỞ ĐẦU

## 1. TÍNH CẤP THIẾT CỦA ĐỀ TÀI

Đáp ứng nhu cầu thông tin của người dùng: Với sự phát triển của công nghệ thông tin, người dùng ngày càng tìm kiếm thông tin trực tuyến để tìm hiểu về giá cả và xu hướng bất động sản. Dự đoán giá nhà và chung cư từ dữ liệu từ nhatot đáp ứng nhu cầu này bằng cách cung cấp thông tin chi tiết và chính xác về giá cả, giúp người dùng có cái nhìn rõ ràng và đáng tin cậy về thị trường bất động sản tại thành phố Hồ Chí Minh.

Đóng góp vào lĩnh vực nghiên cứu bất động sản: Việc dự đoán giá nhà và chung cư tại thành phố Hồ Chí Minh có ý nghĩa quan trọng trong lĩnh vực bất động sản. Đề tài cung cấp phương pháp và kỹ thuật dự đoán giá cả nhanh chóng và chính xác, đáng tin cậy cho các nhà đầu tư, nhà phát triển và người mua bất động sản. Điều này giúp cải thiện quyết định đầu tư và giao dịch bất động sản, tạo ra lợi ích kinh tế và phát triển bền vững cho thành phố Hồ Chí Minh.

Tiềm năng phát triển và ứng dụng cho các nghiên cứu tương lai: Đề tài cung cấp một cơ sở nền tảng cho các nghiên cứu liên quan đến dự đoán giá nhà và chung cư. Các phương pháp và kỹ thuật được áp dụng trong đề tài có thể được mở rộng và tùy chỉnh cho các lĩnh vực và thị trường bất động sản khác. Điều này mở ra tiềm năng phát triển cho các nghiên cứu tương lai trong lĩnh vực dự đoán giá nhà và chung cư, góp phần nâng cao hiệu quả và chính xác của các phương pháp dự đoán.

## 2. MỤC ĐÍCH CỦA ĐỀ TÀI

Nắm bắt xu hướng và biến động giá cả: Mục đích chính của đề tài là phân tích và dự đoán giá nhà và chung cư tại thành phố Hồ Chí Minh dựa trên dữ liệu từ nhatot. Qua đó, nghiên cứu nhằm mục đích cung cấp thông tin cụ thể về xu hướng và biến động giá cả trong thị trường bất động sản. Điều này giúp người đầu tư và người mua bất động sản có cái nhìn tổng quan về tình hình thị trường và đưa ra quyết định mua bán dựa trên thông tin đáng tin cậy.

Cải thiện quyết định đầu tư và giao dịch bất động sản: Bằng cách áp dụng các phương pháp dự đoán giá cả, đề tài nhằm cung cấp một công cụ hữu ích để đánh giá và dự báo giá trị tài sản bất động sản tại thành phố Hồ Chí Minh. Điều này giúp cải thiện quyết định đầu tư và giao dịch của các nhà đầu tư và người mua, từ đó tối ưu hóa lợi nhuận và giảm rủi ro trong các hoạt động liên quan đến bất động sản.

Đóng góp vào phát triển bền vững của thị trường bất động sản: Việc dự đoán giá nhà và chung cư tại thành phố Hồ Chí Minh từ dữ liệu từ nhatot có ý nghĩa quan trọng trong việc xác định xu hướng và tăng trưởng của thị trường bất động sản. Đề tài đóng góp vào việc hiểu rõ hơn về cung cầu, yếu tố ảnh hưởng đến giá cả, và sự biến động của thị trường bất động sản. Điều này hỗ trợ cho quy hoạch và phát triển bền vững của thành phố Hồ Chí Minh, từ việc quản lý hợp lý cung cầu bất động sản đến đáp ứng nhu cầu nhà ở của người dân.

## 3. CÁCH TIẾP CẬN VÀ PHƯƠNG PHÁP NGHIÊN CỨU

### 3.1. Đối tượng nghiên cứu

Đối tượng chính của nghiên cứu là thị trường bất động sản thành phố Hồ Chí Minh. Nghiên cứu sẽ tập trung vào việc dự đoán giá nhà và chung cư, sử dụng dữ liệu từ nhatot. Các dữ liệu về giá cả, diện tích, vị trí địa lý, tiện ích xung quanh và các yếu tố khác liên quan sẽ được thu thập và phân tích để xây dựng mô hình dự đoán.

### 3.2. Phạm vi nghiên cứu

Phạm vi của nghiên cứu sẽ tập trung vào thành phố Hồ Chí Minh, một thị trường bất động sản lớn và phát triển tại Việt Nam. Dữ liệu về giá nhà và chung cư sẽ được thu thập từ nhatot, một nguồn thông tin uy tín về thị trường bất động sản. Nghiên cứu sẽ tập trung vào việc xây dựng mô hình dự đoán giá cả dựa trên các yếu tố quan trọng như diện tích, vị trí địa lý, tiện ích và các yếu tố khác có thể ảnh hưởng đến giá trị tài sản bất động sản.

### 3.3. Cách tiếp cận và phương pháp nghiên cứu

Để dự đoán giá nhà và chung cư thành phố Hồ Chí Minh, nghiên cứu sẽ sử dụng các phương pháp và kỹ thuật trong lĩnh vực học máy và khai phá dữ liệu. Cụ thể, một phương pháp dự đoán có thể là sử dụng mô hình hồi quy, trong đó dữ liệu về các yếu tố như diện tích, vị trí địa lý và các tiện ích xung quanh được sử dụng để đưa ra dự đoán về giá cả bất động sản.

Ngoài ra, để cải thiện độ chính xác của mô hình dự đoán, nghiên cứu cũng có thể sử dụng các phương pháp phân tích đa biến và kỹ thuật học máy phức tạp khác như mạng nơ-ron nhân tạo (Artificial Neural Networks), máy vector hỗ trợ (Support Vector Machines) hoặc các thuật toán tăng cường (Boosting algorithms).

Tổng hợp lại, cách tiếp cận và phương pháp nghiên cứu trong đề tài "Dự đoán giá nhà và chung cư thành phố Hồ Chí Minh với dữ liệu từ nhatot" sẽ sử dụng các phương pháp học máy và khai phá dữ liệu để xây dựng mô hình dự đoán giá cả bất động sản. Nghiên cứu sẽ tập trung vào thành phố Hồ Chí Minh và sử dụng dữ liệu từ nhatot làm nguồn thông tin chính.

## 4. KẾT QUẢ VÀ DỰ KIẾN ĐẠT ĐƯỢC

Kết quả dự kiến của đề tài "Dự đoán giá nhà và chung cư thành phố Hồ Chí Minh với dữ liệu từ nhatot" là xây dựng mô hình dự đoán giá cả chính xác, cung cấp thông tin hữu ích cho quyết định đầu tư và đóng góp vào phát triển bền vững của thị trường bất động sản.

## 5. BỐ CỤC BÀI BÁO CÁO

# PHẦN NỘI DUNG

## CHƯƠNG 1: CƠ SỞ LÝ THUYẾT VÀ CÁC NGHIÊN CỨU LIÊN QUAN

### 1.1. Những vấn đề cơ bản về bất động sản

#### 1.1.1. Khái niệm bất động sản

Bất động sản là loại hình tài sản có giá trị kinh tế và có tính định hình cao. Bất động sản bao gồm tất cả các loại tài sản vật chất có thể được sở hữu như đất đai, nhà cửa, căn hộ chung cư, trung tâm thương mại, khu công nghiệp, khu nghỉ dưỡng...

Điểm đặc trưng của bất động sản là sự cố định về vị trí, không thể di chuyển được theo chủ sở hữu. Đất đai, nhà cửa luôn nằm tại một địa chỉ cụ thể. Giá trị của bất động sản phụ thuộc vào nhiều yếu tố như vị trí địa lý, quy mô, chất lượng xây dựng, tiện ích xung quanh, cơ sở hạ tầng...

Bất động sản là một loại tài sản quan trọng có thể cho thuê, thế chấp, mua bán để sinh lời. Ngành bất động sản cung cấp nhiều dịch vụ liên quan đến phát triển dự án, môi giới, quản lý, vận hành...

#### 1.1.2. Phân loại bất động sản

Bất động sản là một loại tài sản quan trọng, có giá trị lớn và đa dạng về loại hình. Việc phân loại bất động sản giúp hiểu rõ hơn các loại hình bất động sản cũng như đặc điểm riêng của từng loại. Bất động sản có thể được phân loại theo nhiều tiêu chí sau:

* Phân loại theo loại hình xây dựng
* Nhà phố, biệt thự, liền kề: Loại bất động sản nhà ở độc lập dành cho một hộ gia đình.
* Chung cư: Loại hình bất động sản xây dựng cao tầng với nhiều căn hộ trong một tòa nhà.
* Khu đô thị, khu căn hộ: Quy hoạch phức hợp nhiều chức năng trong cùng một khu vực.
* Phân loại theo mục đích sử dụng
* Nhà ở: Gồm các loại nhà riêng lẻ, chung cư phục vụ nhu cầu ở.
* Thương mại dịch vụ: Trung tâm thương mại, trung tâm giải trí, khách sạn...
* Công nghiệp: Khu công nghiệp, nhà xưởng, kho bãi
* Văn phòng: Toà nhà văn phòng, tòa nhà hành chính.
* Các tiêu chí phân loại khác
* Theo khu vực địa lý: Nội đô, ngoại ô, ven biển, miền núi...
* Theo chủ sở hữu: Nhà nước, doanh nghiệp, cá nhân.
* Theo mức độ sang trọng: Cao cấp, trung cấp, bình dân.

## CHƯƠNG 2: TỔNG QUAN VỀ THỊ TRƯỜNG BẤT ĐỘNG SẢN TẠI THÀNH PHỐ HỒ CHÍ MINH

### 2.1. Giới thiệu

Thị trường bất động sản Thành phố Hồ Chí Minh đang dần phục hồi sau ảnh hưởng của đại dịch Covid-19

Theo nhận định của các chuyên gia ngành bất động sản, thị trường nhà đất tại TP. Hồ Chí Minh đang dần có dấu hiệu hồi phục sau khoảng thời gian trầm lắng do tác động của đại dịch. Cụ thể, giá bất động sản đang dần ổn định trở lại sau những biến động mạnh trước đây. Đồng thời, lượng giao dịch mua bán, chuyển nhượng cũng ghi nhận xu hướng tăng trở lại so với thời điểm cao điểm của dịch bệnh.

Nguyên nhân được cho là do Chính phủ đã triển khai nhiều chính sách hỗ trợ thiết thực nhằm thúc đẩy phục hồi kinh tế và thị trường bất động sản. Cụ thể, nhiều dự án bất động sản lớn đã được gỡ khó về pháp lý và được phép huy động vốn từ nhà đầu tư. Đồng thời, lãi suất huy động vốn tại các ngân hàng thương mại cũng giảm mạnh, tạo đà cho nhu cầu mua sắm nhà đất trở lại.

Tuy nhiên, tiêu thụ sản phẩm vẫn chưa được như kỳ vọng do sức mua của người dân bị suy giảm sau đại dịch. Dự báo trong thời gian tới, khi lãi suất giảm thấp hơn nữa và tình hình dịch bệnh tiếp tục được kiểm soát, thị trường bất động sản có thể phục hồi mạnh mẽ hơn.

### 2.2 Phát triển kinh tế và tăng trưởng dân số

### 2.3. Tình hình cung cầu bất động sản

Về tình hình cung cầu trên thị trường bất động sản TP. HCM trong 9 tháng đầu năm 2023:

Về nguồn cung, theo khảo sát của Savills Việt Nam, trong 9 tháng đầu năm 2023, tổng lượng sản phẩm mới đưa ra thị trường ước đạt khoảng 30.000 căn (Savills Việt Nam, 2023). So với cùng kỳ năm ngoái, nguồn cung bất động sản tăng khoảng 5-10%. Tuy nhiên, phần lớn là các dự án đang triển khai từ trước tiếp tục đưa ra thị trường, số lượng dự án mới được phát triển còn hạn chế.

Về nhu cầu, theo báo cáo thị trường quý III/2023 của DKRA Vietnam, lượng giao dịch bất động sản trong 9 tháng đầu năm đạt khoảng 25.000 giao dịch, tăng 15% so với cùng kỳ năm trước (DKRA Vietnam, 2023). Nhu cầu tập trung chủ yếu vào phân khúc căn hộ trung và bình dân có giá hợp lý tại khu vực trung tâm.

Từ đó, có thể thấy thị trường đang có tình trạng cung vượt cầu ở phân khúc cao cấp, trong khi phân khúc trung và bình dân vẫn còn thiếu hụt nguồn cung để đáp ứng nhu cầu ngày một tăng cao của người dân. Giá bán bất động sản không có biến động đáng kể.

### 2.4. Các yếu tố ảnh hưởng đến giá nhà và chung cư

Giá bất động sản tại TP. Hồ Chí Minh bị ảnh hưởng bởi nhiều yếu tố như:

* Vị trí dự án: Theo nghiên cứu của Jones Lang LaSalle (2023), giá bất động sản ở khu vực trung tâm, gần khu công nghiệp thường cao hơn so với khu vực ngoại ô hoặc vùng ven.
* Trạng thái pháp lý: Có nghiên cứu cho thấy, bất động sản có sổ đỏ, hồ sơ pháp lý rõ ràng sẽ được định giá cao hơn (Savills Việt Nam, 2023).
* Cơ sở hạ tầng và tiện ích: Theo khảo sát của DKRA Vietnam (2022), dự án có đầy đủ tiện ích xung quanh thường cao hơn so với nơi thiếu tiện ích.
* Chất lượng sản phẩm: Căn hộ cao cấp, diện tích rộng, tiêu chuẩn xây dựng tốt sẽ có giá bán cao hơn (Cushman & Wakefield, 2023).
* Tình hình kinh tế vĩ mô: Biến động về lạm phát, lãi suất, thu nhập dân cư đều ảnh hưởng đến giá bất động sản (Sở Xây dựng TPHCM, 2023)

## CHƯƠNG 3. CÁC PHƯƠNG PHÁP DỰ ĐOÁN GIÁ NHÀ VÀ CHUNG CƯ

### 3.1. Phân tích hồi quy tuyến tính

#### 3.1.1. Khái niệm và cơ sở lý thuyết

Phân tích hồi quy tuyến tính là phương pháp định lượng mô tả mối quan hệ giữa biến phụ thuộc với một hoặc nhiều biến độc lập, trong đó giả định mối quan hệ là tuyến tính. Biến phụ thuộc gọi là biến đáp ứng Y, các biến độc lập gọi là biến giải thích X.

Mô hình hồi qui tuyến tính đơn biến dùng để xem xét mối quan hệ tuyến tính giữa biến phụ thuộc y (biến kết cục) và biến độc lập x (biến dự đoán). Phương trình tuyến tính (đường thẳng) đơn biến có dạng:

**y= α + βxi+εi**

Trong đó α là điểm cắt trên trục tung, β là độ dốc (trong thống kê gọi là hệ số hồi qui) và ε là phần dư. Ví dụ tìm sự liên hệ giữa diện tích (x) và giá nhà (y).

Tuy nhiên trong thực tế, thường có rất nhiều yếu tố khác khác nhau dẫn đến một hiện tượng. Ví dụ giá nhà không chỉ do diện tích tác động đến mà bên cạnh đó còn có các yếu tố khác như số phòng ngủ, số phòng tắm, số tầng...

Do đó, ta có mô hình qui tuyến tính đa biến có dạng:

**y= α + β1x1i+ β2x2i +β3x3i + ......+........+ βkxki + εi**

Với ví dụ trên ta có thể viết phương trình hồi quy tuyến tính đa biến với 4 yếu tố (x1 = diện tích, x2 = số phòng ngủ, x3 = số phòng tắm, x4 = số tầng) như sau:

Giá nhà = α + β1(diện tích) + β2(số phòng ngủ) + β3(số phòng tắm) + β4(số tầng) + ε

Các giá trị hệ số hồi qui β (regression coefficient): β0, β1, β2, …, βk được chọn bằng phương pháp bình phương tối thiểu, giúp tối thiểu hóa tổng bình phương phần dư (RSS), yếu tố nào có β cao thì ảnh hưởng nhiều hơn.

**RSS = Σ(yi – ŷi)2**

Trong đó:

* Σ: Ký hiệu tiếng Hy Lạp có nghĩa là tổng
* yi: Giá trị phản hồi thực tế cho quan sát thứ i
* ŷi: Giá trị phản hồi dự đoán dựa trên mô hình hồi quy tuyến tính bội

Trong phân tích hồi qui tuyến tính đa biến, ta cần biết mức độ ảnh hưởng của từng yếu tố lên biến kết cục y (giá nhà trong ví dụ này). Muốn biết mức độ ảnh hưởng cần lưu ý đến các trị số sau:

Hệ số tương quan R (coefficient of correlation): yếu tố nào có R càng lớn thì ảnh hưởng càng nhiều.

Bình phương của R (R square): Đây được biết đến là hệ số xác định. Đó là tỷ lệ phần trăm phương sai trong biến phụ thuộc mà có thể được giải thích bởi các biến giải thích. yếu tố nào có R2 càng lớn thì mối quan hệ giữa yếu tố đó và biến y càng chặt chẽ.

Cách xác định R2:

Đầu tiên, ta cần tính tổng bình phương sai số (SSE) của mô hình hồi quy tuyến tính bằng cách tính tổng bình phương của sai số (chênh lệch giữa giá trị dự đoán và giá trị thực tế) cho tất cả các điểm dữ liệu trong tập huấn luyện:

**SSE = Σ(y - ŷ)²**

Trong đó:

* y là giá trị thực tế của biến phụ thuộc.
* ŷ là giá trị dự đoán của biến phụ thuộc dựa trên các giá trị độc lập được sử dụng trong mô hình.

Tiếp đến, tính tổng bình phương sai số của trung bình (SST) bằng cách tính tổng bình phương chênh lệch giữa giá trị thực tế và giá trị trung bình của biến phụ thuộc cho tất cả các điểm dữ liệu trong tập huấn luyện:

**SST = Σ(y - ȳ)²**

Trong đó:

* ȳ là giá trị trung bình của biến phụ thuộc.

Tính hệ số xác định (R²) bằng cách tính tỷ lệ phần trăm giải thích được của biến phụ thuộc bởi các biến độc lập trong mô hình:

**R² = 1 - (SSE / SST)**

Trị số p (p value): càng nhỏ mức ảnh hưởng càng mạnh.

#### 3.1.2. Các giả định của mô hình hồi quy tuyến tính

Phần này giới thiệu các giả định cơ bản của mô hình hồi quy tuyến tính và vai trò quan trọng của chúng đối với việc ước lượng mô hình. Việc hiểu rõ các giả định này sẽ giúp đánh giá được tính hợp lệ của các kết quả ước lượng. Nếu mô hình không đáp ứng đủ các giả định, các hệ số ước lượng có thể bị sai lệch hoặc không nhất quán dẫn đến kết quả phân tích không đáng tin cậy.

Đầu tiên là các giả định cơ bản đảm bảo sự tồn tại của hệ số hồi quy tuyến tính:

* Hệ số hồi quy phải ở dạng hằng số: Nếu ở dạng số mũ, log, tỷ số sẽ không phải là một mối quan hệ tuyến tính đơn giản giữa biến phụ thuộc và độc lập.
* Các biến độc lập là cố định hoặc phi ngẫu nhiên: Giá trị các biến độc lập không thay đổi khi lấy mẫu khác nhau để có thể phân tích được ảnh hưởng của chúng.
* Các biến độc lập có phương sai lớn hơn 0: Có sự khác biệt giá trị quan sát để quan sát được mối quan hệ.

Tiếp đến là các giả định đảm bảo cho hệ số hồi quy không bị sai lệch, nhất quán và hiệu quả:

* Giá trị trung bình của sai số bằng 0: Loại trừ ảnh hưởng các yếu tố khác.
* Phương sai của sai số là một hằng số: Ảnh hưởng của biến độc lập lên phụ thuộc là như nhau.
* Các sai số không tự tương quan: Loại trừ ảnh hưởng thời gian.
* Các biến độc lập không có quan hệ tuyến tính hoàn hảo: Tránh đa cộng tuyến.
* Mô hình được thiết kế đúng: Đảm bảo không bỏ sót biến quan trọng.

### B. Mô hình dự báo thống kê

### C. Mạng nơ-ron nhân tạo

### D. Các phương pháp khác

## CHƯƠNG 4. THU THẬP DỮ LIỆU VÀ XỬ LÝ DỮ LIỆU

### 4.1. Nguồn dữ liệu

Website https://www.nhatot.com là một trong những nguồn thông tin uy tín về thị trường bất động sản tại TP. Hồ Chí Minh. Trang web liên tục cập nhật các tin rao bán, cho thuê căn hộ chung cư, nhà riêng do chủ đầu tư, công ty môi giới và cá nhân đăng tải.

Các thông tin được cung cấp kèm theo mỗi tin rao bán bao gồm: địa chỉ, phường quận cụ thể; diện tích sàn, số phòng ngủ, toilet; năm xây dựng; giá bán/giá thuê; ảnh minh họa sản phẩm. Đây là những thông tin cần thiết phục vụ mục đích nghiên cứu thị trường và phân tích giá.

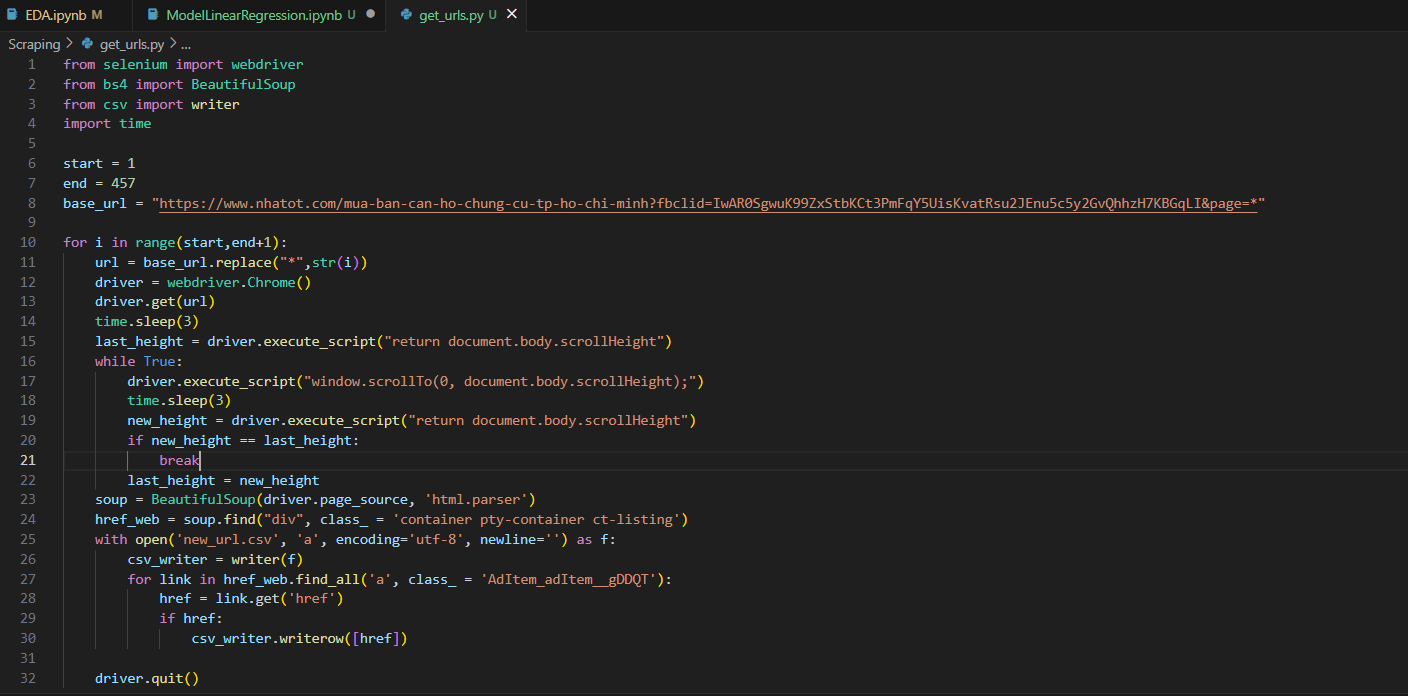
Lượng thông tin trên website được cập nhật hàng ngày, số lượng tin rao bán lớn và phủ khắp các quận huyện TP.HCM. Đặc biệt, website cung cấp nhiều loại hình bất động sản khác nhau như chung cư, nhà phố, biệt thự...cho phép phân tích đa chiều.

### 4.2. Phương pháp thu thập dữ liệu

Để tiến hành nghiên cứu và phân tích thị trường bất động sản, bước đầu tiên là thu thập dữ liệu từ các nguồn tin cậy. Một trong những nguồn dữ liệu quan trọng là website https://www.nhatot.com cung cấp nhiều thông tin chi tiết về căn hộ chung cư rao bán tại TPHCM.

Để thu thập toàn bộ dữ liệu trên website này, em đã xây dựng phần mềm crawl dữ liệu bằng cách sử dụng ngôn ngữ lập trình Python có tên là get\_urls.py. Cụ thể, chương trình sẽ thực hiện các bước sau:

* Truy cập từng trang trên website và lấy nội dung bằng thư viện Selenium.
* Sử dụng Beautiful Soup phân tích cấu trúc DOM của trang và trích xuất dữ liệu.
* Lặp lại quá trình trên cho tất cả các trang có tin rao bán.
* Lưu trữ dữ liệu đã thu thập vào cơ sở dữ liệu để quản lý và phân tích tiếp theo.



Mục đích chính của đoạn code là thu thập danh sách các tin rao bán bất động sản trên website nhà đất. Đây là nguồn dữ liệu quan trọng cần phải khai thác để phục vụ nghiên cứu thị trường.

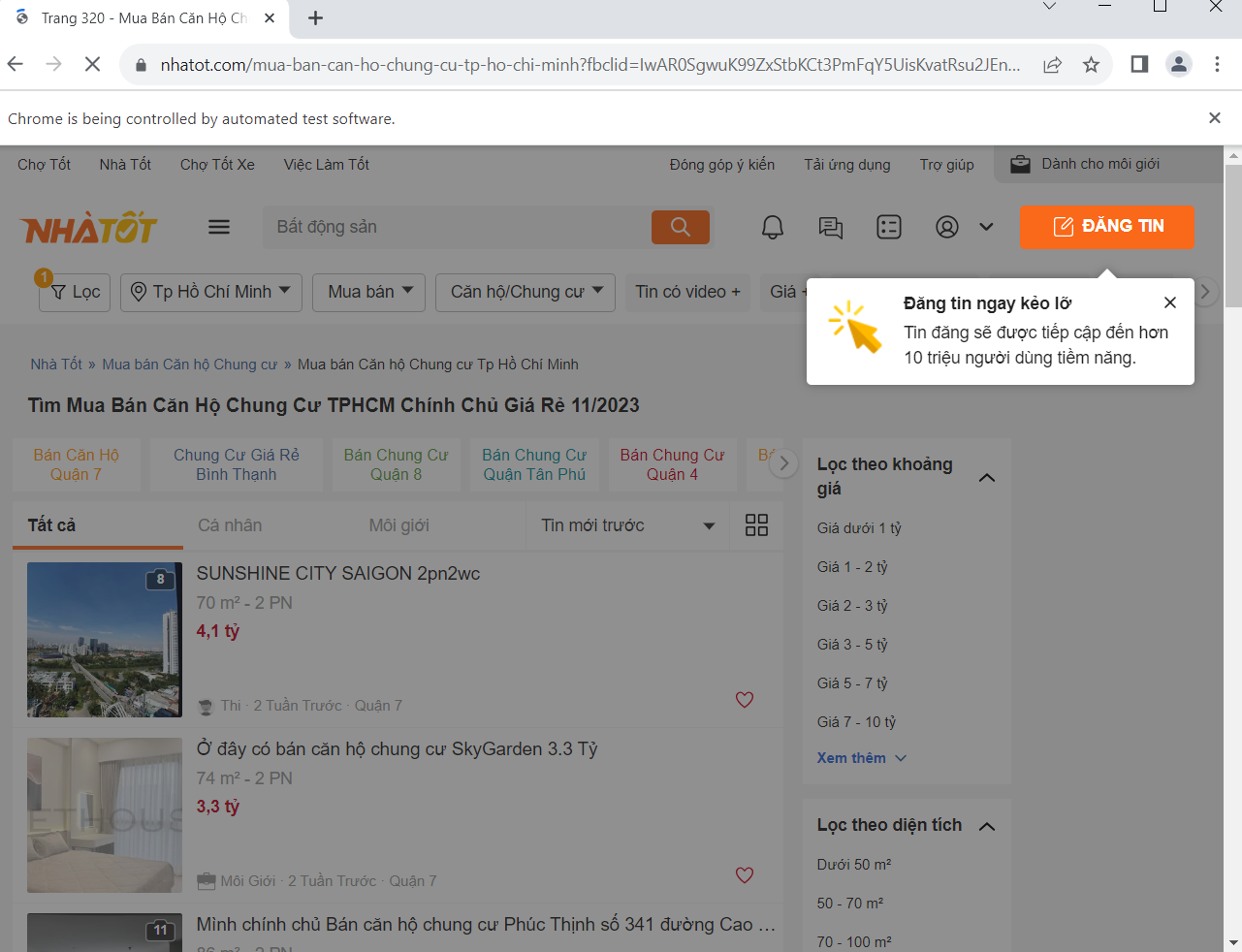
Tuy nhiên, với số lượng tin lớn trên nhiều trang, việc thu thập thủ công sẽ mất nhiều thời gian. Do đó, chương trình sử dụng công nghệ crawl tự động để lấy dữ liệu hiệu quả hơn.

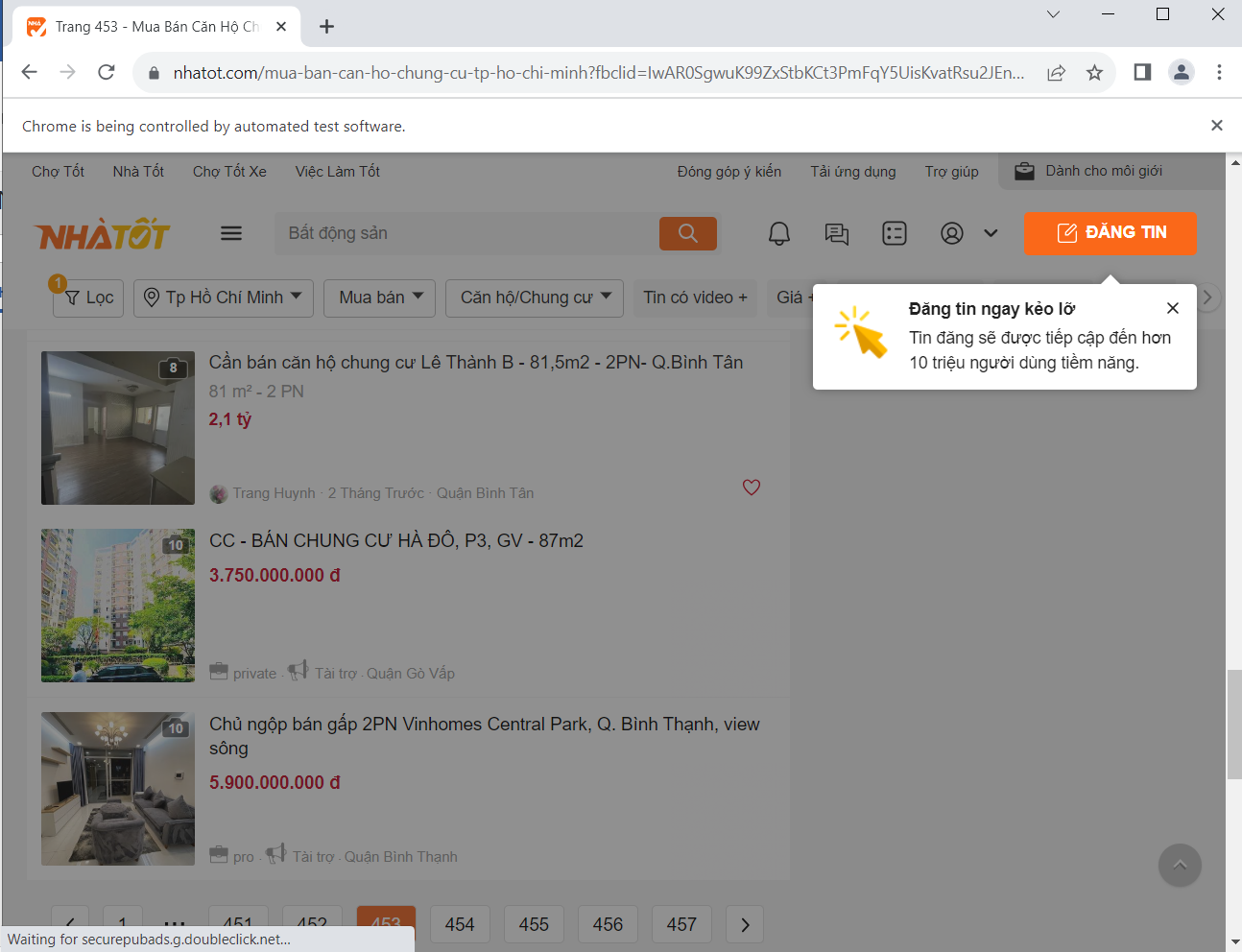
Cụ thể, Selenium được dùng điều khiển trình duyệt, truy cập từng trang theo quy tắc định nghĩa. Beautiful Soup phân tích DOM tree, tìm kiếm và lấy thông tin link các tin rao bán.

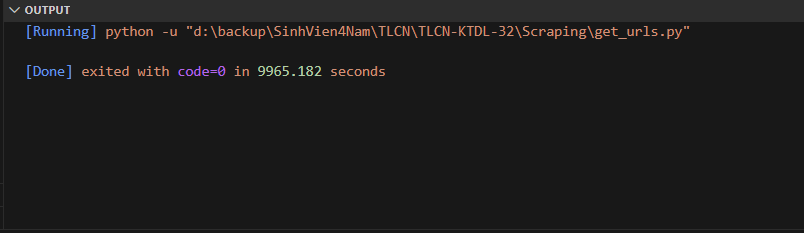
Sau khi thu thập xong, dữ liệu được lưu vào file csv để quản lý và sẵn sàng cho các bước xử lý, phân tích tiếp theo phục vụ nghiên cứu.

Bước thu thập dữ liệu này góp phần cung cấp nguồn dữ liệu số lượng lớn cho quá trình nghiên cứu và rút ra kết luận cuối cùng. Từ đó, mục tiêu đề ra có thể đạt được.

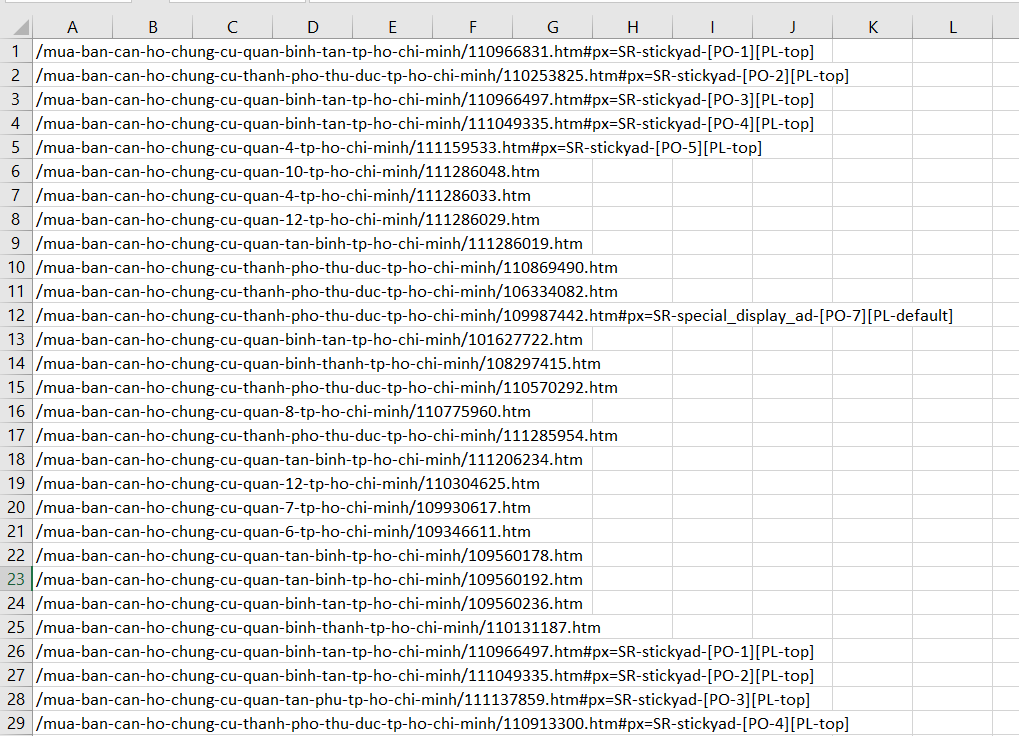
Dưới đây là hình ảnh của quá trình việc thu thập các url từ đoạn mã get\_urls.py.



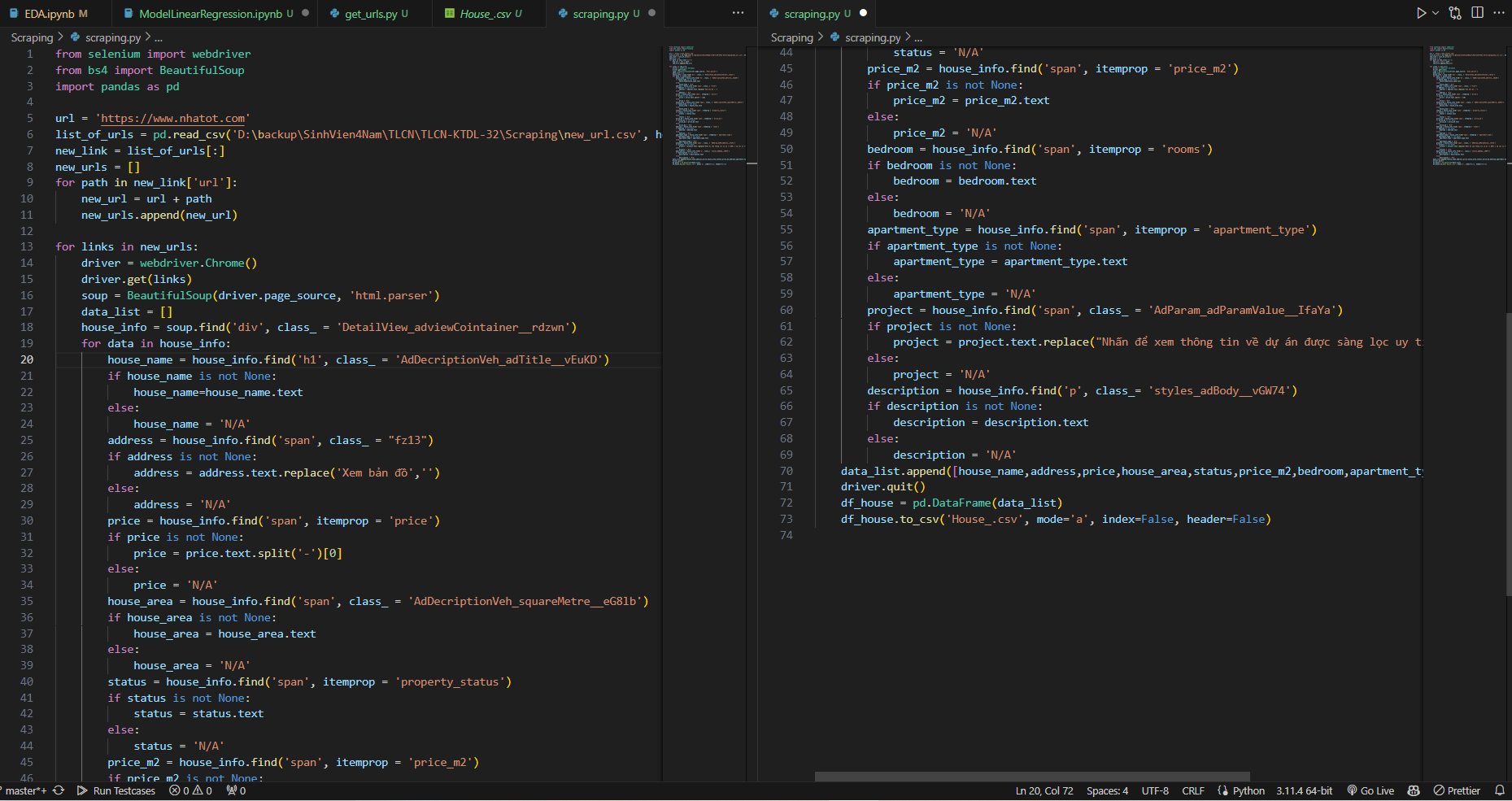




Dữ liệu đường dẫn sau khi thu thập được



File scraping.py



Đoạn code trên là một đoạn mã Python được sử dụng để tự động thu thập thông tin về các nhà ở từ một danh sách các URL và lưu kết quả vào một file CSV.

Trước tiên, các thư viện cần thiết như Selenium, BeautifulSoup và Pandas được import vào chương trình. Sau đó, biến url được khai báo để lưu trữ đường dẫn gốc của trang web mà chúng ta muốn thu thập dữ liệu.

Tiếp theo, chúng ta đọc danh sách các URL từ một file CSV và tạo một danh sách mới để chứa các URL này. URL gốc cũng được thêm vào danh sách mới.

Tiếp theo là một vòng lặp, trong đó chúng ta sử dụng Selenium để khởi tạo một trình duyệt web (ở đây là Chrome) và truy cập vào từng URL trong danh sách mới.

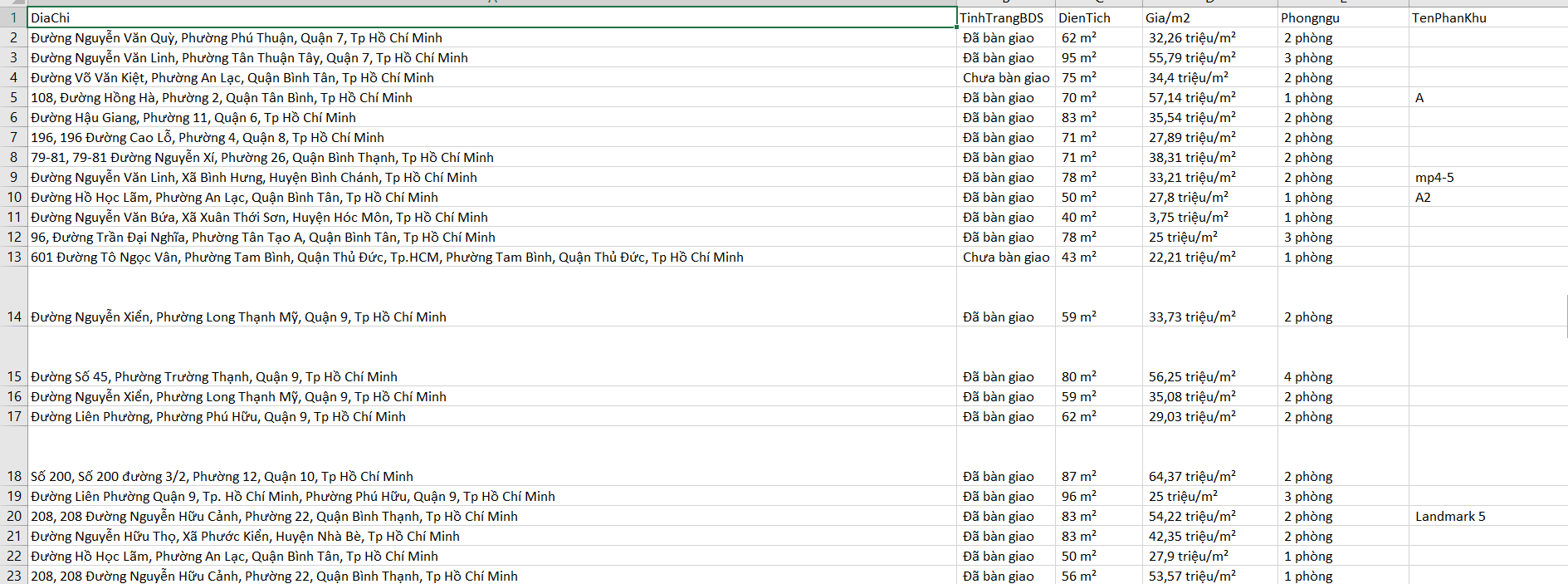
Sau khi trang web được tải, chúng ta sử dụng BeautifulSoup để phân tích cú pháp HTML và lấy thông tin về các nhà ở từ trang web này. Thông tin như tên nhà, địa chỉ, giá, diện tích, tình trạng, giá mỗi mét vuông, số phòng ngủ, loại căn hộ, dự án và mô tả được lấy ra và lưu vào một danh sách.

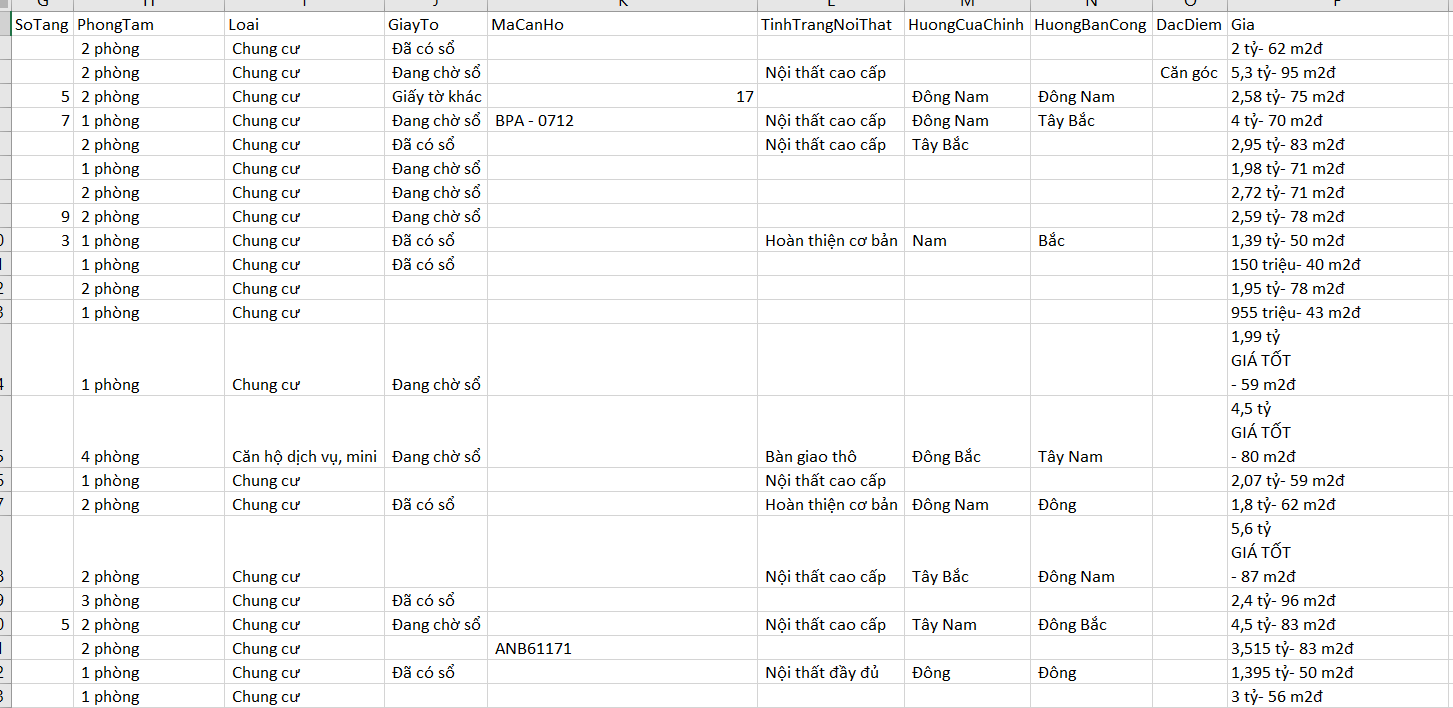
Sau khi thu thập thông tin từ một trang web, chúng ta đóng trình duyệt và tiếp tục vòng lặp để xử lý các trang web tiếp theo trong danh sách.

Cuối cùng, chúng ta tạo một DataFrame từ danh sách thông tin nhà và ghi dữ liệu vào một file CSV để lưu trữ. Điều này cho phép chúng ta dễ dàng truy cập và sử dụng thông tin về các nhà ở thu thập được sau này.

Tổng thể, đoạn code này cung cấp một phương pháp tự động và hiệu quả để thu thập thông tin về các nhà ở từ một danh sách các URL và tổ chức dữ liệu thu thập được vào một file CSV.

Dưới đây là kết quả sau khi lưu data vào raw\_data.csv được từ các urls:



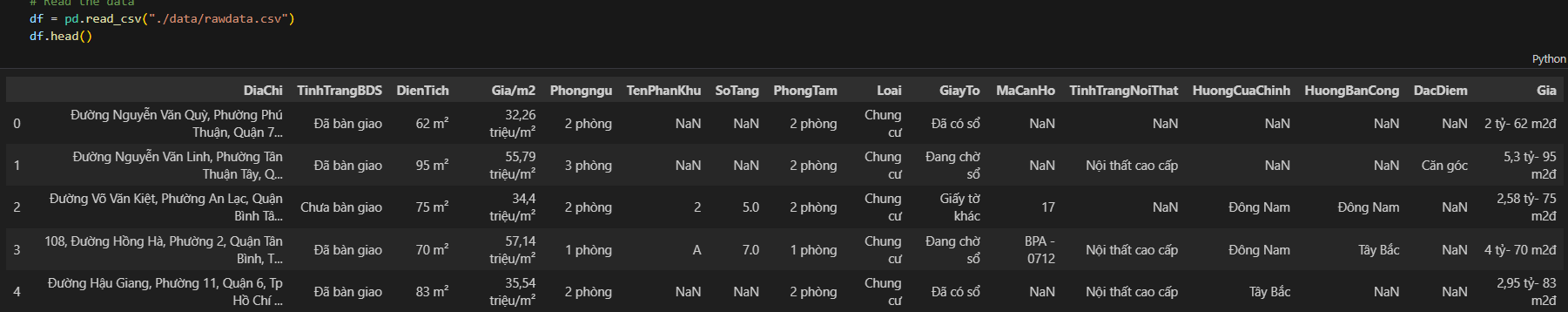


Mô tả về các biến có trong tập dữ liệu:

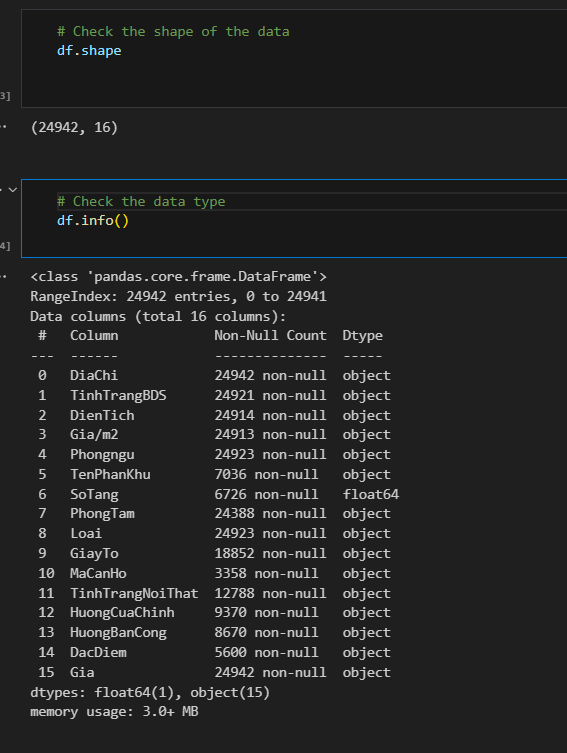
* Địa chỉ (DiaChi): Đây là trường chứa địa chỉ của bất động sản, tức là vị trí nơi nó được đặt.
* Tình trạng bất động sản (TinhTrangBDS): Đây là trường mô tả tình trạng hiện tại của bất động sản, chẳng hạn như "Đã bàn giao", "Chưa bàn giao" ...
* Diện tích (DienTich): Đây là trường chứa diện tích của bất động sản, được đo bằng đơn vị mét vuông.
* Giá/m2 (Gia/m2): Đây là trường chứa giá bán hoặc giá thuê của bất động sản mỗi mét vuông.
* Số phòng ngủ (Phongngu): Đây là trường mô tả số lượng phòng ngủ có trong bất động sản.
* Tên phân khu (TenPhanKhu): Đây là trường chứa tên phân khu hoặc tên dự án của bất động sản.
* Số tầng (SoTang): Đây là trường mô tả số tầng của bất động sản.
* Phòng tắm (PhongTam): Đây là trường mô tả số lượng phòng tắm có trong bất động sản.
* Loại (Loai): Đây là trường mô tả loại bất động sản, chẳng hạn như "Chung cư", "căn hộ dịch vụ, mini", "penthouse", "duplex", "quần thể", "cư xá" và các loại khác.
* Giấy tờ (GiayTo): Đây là trường mô tả tình trạng giấy tờ pháp lý của bất động sản, ví dụ như "đã có sổ", "đang chờ sổ", "giấy tờ khác" và các tình trạng tương tự.
* Mã căn hộ (MaCanHo): Đây là trường chứa mã căn hộ hoặc mã định danh duy nhất cho mỗi bất động sản.
* Tình trạng nội thất (TinhTrangNoiThat): Đây là trường mô tả tình trạng nội thất của bất động sản, chẳng hạn như "hoàn thiện cơ bản", "nội thất cao cấp", "nội thất đầy đủ", "bàn giao thô" và các tình trạng tương tự.
* Hướng cửa chính (HuongCuaChinh): Đây là trường mô tả hướng của cửa chính của bất động sản, chẳng hạn như "đông", "tây", "nam", "bắc", và các hướng khác.
* Hướng ban công (HuongBanCong): Đây là trường mô tả hướng của ban công của bất động sản, tương tự như trường Hướng cửa chính.
* Đặc điểm (DacDiem): Đây là trường chứa các đặc điểm đặc biệt, thông tin mô tả về bất động sản. Chẳng hạn như là "căn góc" ...
* Giá (Gia): Đây là trường chứa giá bán hoặc giá thuê của bất động sản.

### 4.3. Tiền xử lý dữ liệu

In ra 5 dòng đầu của tập dữ liệu:



Mô tả về tập dữ liệu:



Số lượng null của các biến trong tập dữ liệu:



## CHƯƠNG 4. XÂY DỰNG MÔ HÌNH DỰ ĐOÁN GIÁ NHÀ VÀ CHUNG CƯ

### A. Mô tả mô hình

### B. Huấn luyện mô hình

### C. Đánh giá và tinh chỉnh mô hình

### VI. Dự đoán giá nhà và chung cư tại thành phố Hồ Chí Minh

### A. Kết quả dự đoán

### B. So sánh với giá thực tế

### C. Đánh giá hiệu suất và độ chính xác của mô hình

### VII. Thảo luận và kết luận

### A. Thảo luận kết quả

### B. Đóng góp của nghiên cứu

### C. Hạn chế và hướng phát triển tiếp theo

### VIII. Tài liệu tham khảo

http://xaydung.gov.vn/vn/tin-tuc/1206/76482/thi-truong-bat-dong-san-thanh-pho-ho-chi-minh-co-dau-hieu-phuc-hoi.aspx

Savills Việt Nam, 2023. Báo cáo thị trường quý III/2023.

DKRA Vietnam, 2023. Báo cáo thị trường bất động sản quý III/2023.

Savills Việt Nam, 2023. Báo cáo giá chào bán bất động sản quý III.

DKRA Vietnam, 2022. Yếu tố ảnh hưởng đến giá bất động sản.

Cushman & Wakefield, 2023. Báo cáo thị trường bất động sản quý III.

Sở Xây dựng TPHCM, 2023. Báo cáo tình hình thị trường quý III.