**HỌC VIỆN KỸ THUẬT QUÂN SỰ**

**KHOA CÔNG NGHỆ THÔNG TIN**

**BÁO CÁO MÔN HỌC**

**NHẬP MÔN CƠ SỞ DỮ LIỆU LỚN**

**ĐỀ TÀI: Quản lý nhà đất.**

|  |  |
| --- | --- |
| **Sinh viên thực hiện:** | **Nguyễn Hữu Mạnh** |
|  | **Trần Trung Kiên** |

*Hà Nội, 12/2020*

# **PHÂN CÔNG CÔNG VIỆC**

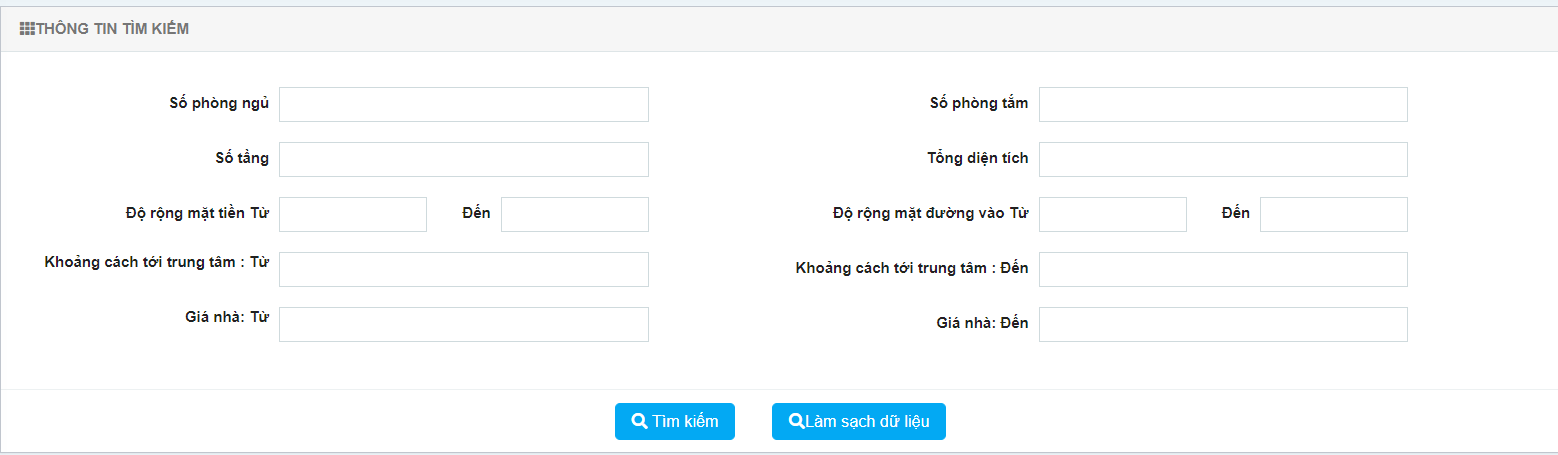
|  |  |
| --- | --- |
| **Công việc** | **Sinh viên thực hiện** |
| Xây dựng chức năng CRUD quản lý thông tin nhà | Trần Trung Kiên, Nguyễn Hữu Mạnh |
| Xây dựng chức năng gợi ý giá nhà có sử dụng học máy | Nguyễn Hữu Mạnh |
| Trực quan hóa dữ liệu | Nguyễn Hữu Mạnh |
| Phân tích chất lượng dữ liệu | Trần Trung Kiên |
| Đề xuất giải pháp làm sạch dữ liệu | Nguyễn Hữu Mạnh |

# 1. Xây dựng các chức năng CRUD

## 1.1. Chức năng tìm kiếm

Chức năng tim kiếm được xây dựng tìm kiếm theo các tiêu chí như: Số phòng ngủ, số phòng tắm,số tầng, tổng diện tích, độ rộng mặt tiền, độ rộng mặt đường vào, khoảng cách tới trung, giá nhà.

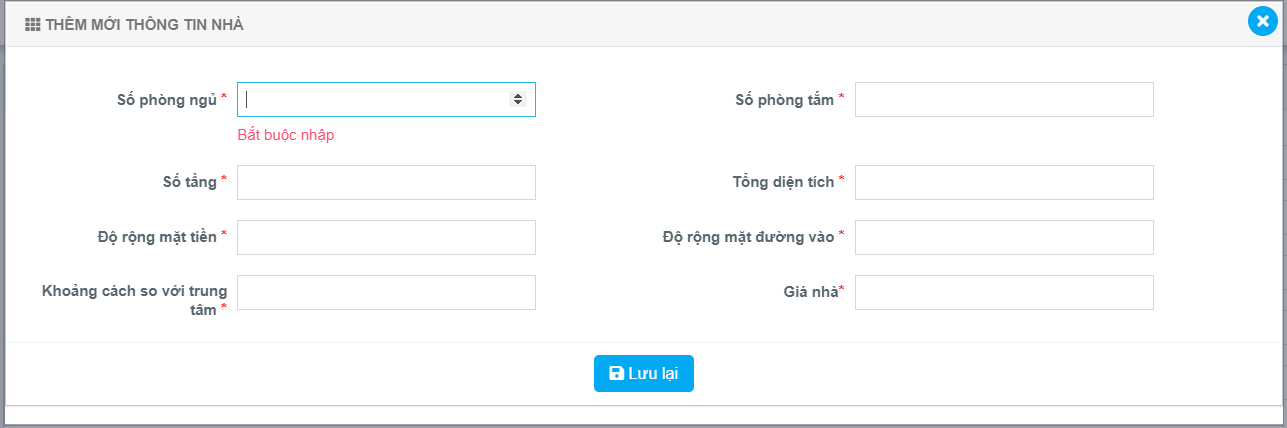
Khi mặc định chưa có thông tin tìm kiếm sẽ tìm kiếm tất cả bản ghi. Danh sách được sắp xếp theo thứ tự giảm dần theo thời gian.

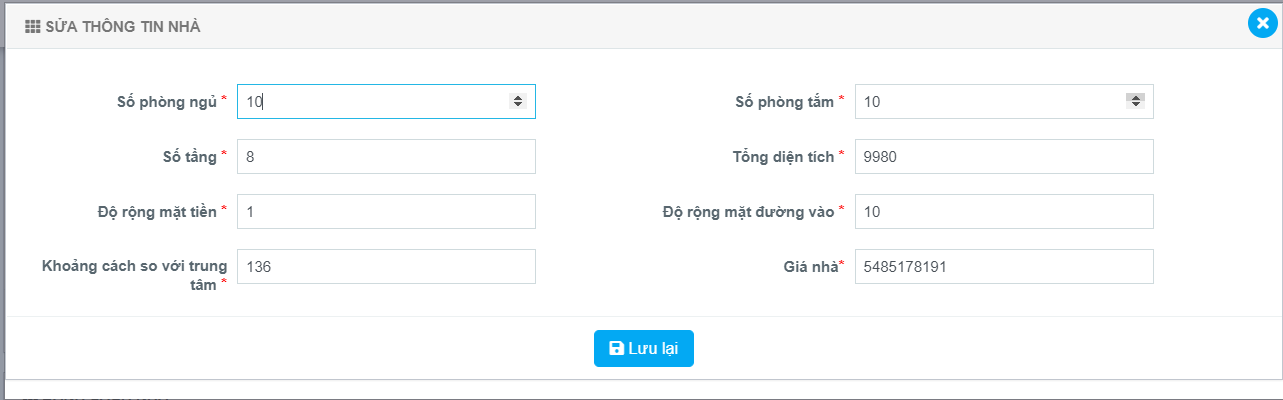






## 1.2. Chức năng thêm mới, sửa





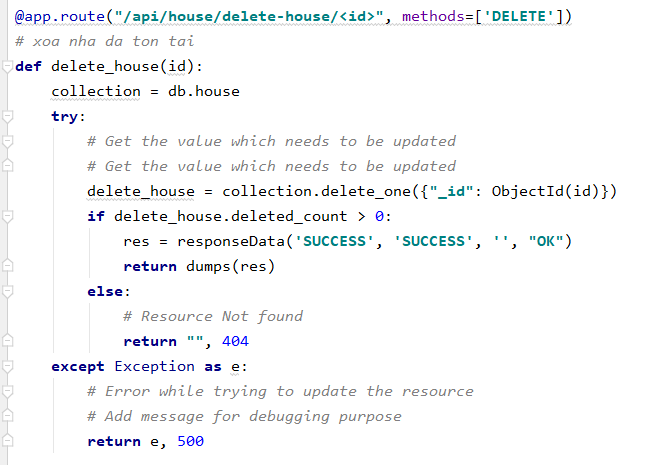
Khi thêm mới, sửa thông tin nhà cần điền đầy đủ các thông tin các trường. Sau khi lưu và sửa , sẽ tự động tìm kiếm lại. Bản ghi mới sẽ hiển thị lên đầu.



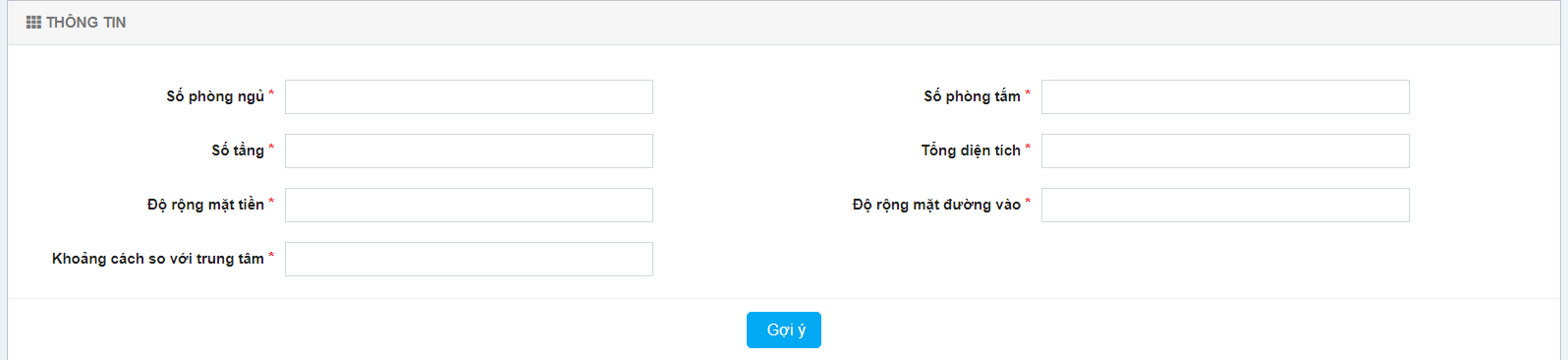


## 1.3. Chức năng xóa.

Chọn xóa bản ghi trên danh sách dữ liệu. Sẽ hỏi người dùng có xóa hay không. Nếu có xóa sẽ thực hiện xóa bản ghi và tìm kiếm lại danh sách.



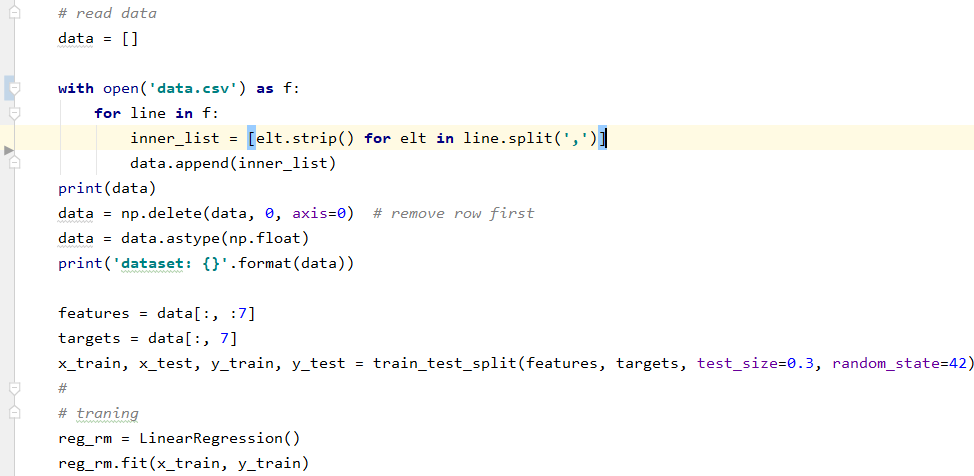
# 2. Xây dựng chức năng gợi ý giá nhà có sử dụng học máy



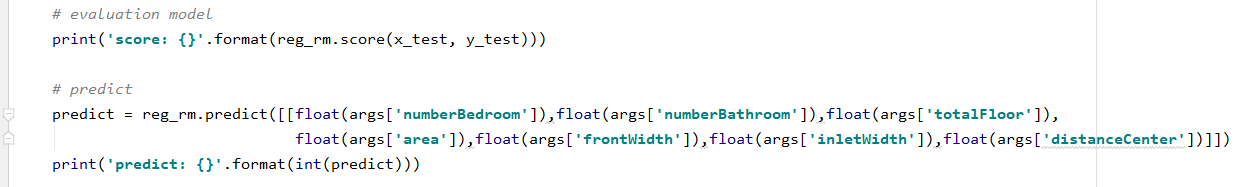
Dữ liệu được sử dụng là dữ liệu lấy từ dữ liệu đang quản lý được export ra file dạng csv

Sử dụng thuật toán LinearRegression để đưa ra gợi ý giá nhà từ các thông tin được nhập trên màn hình.

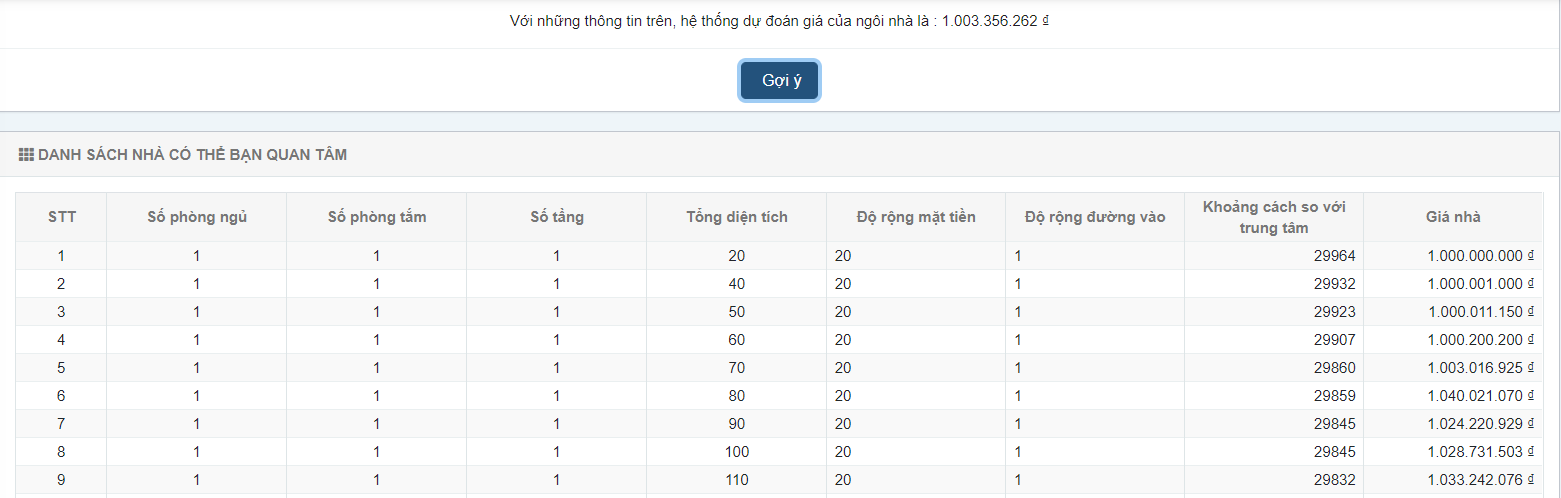
Đọc dữ liệu, chia dữ liệu training và test:



Đưa ra hệ số score và kết quả gợi ý từ đầu vào trên màn hình

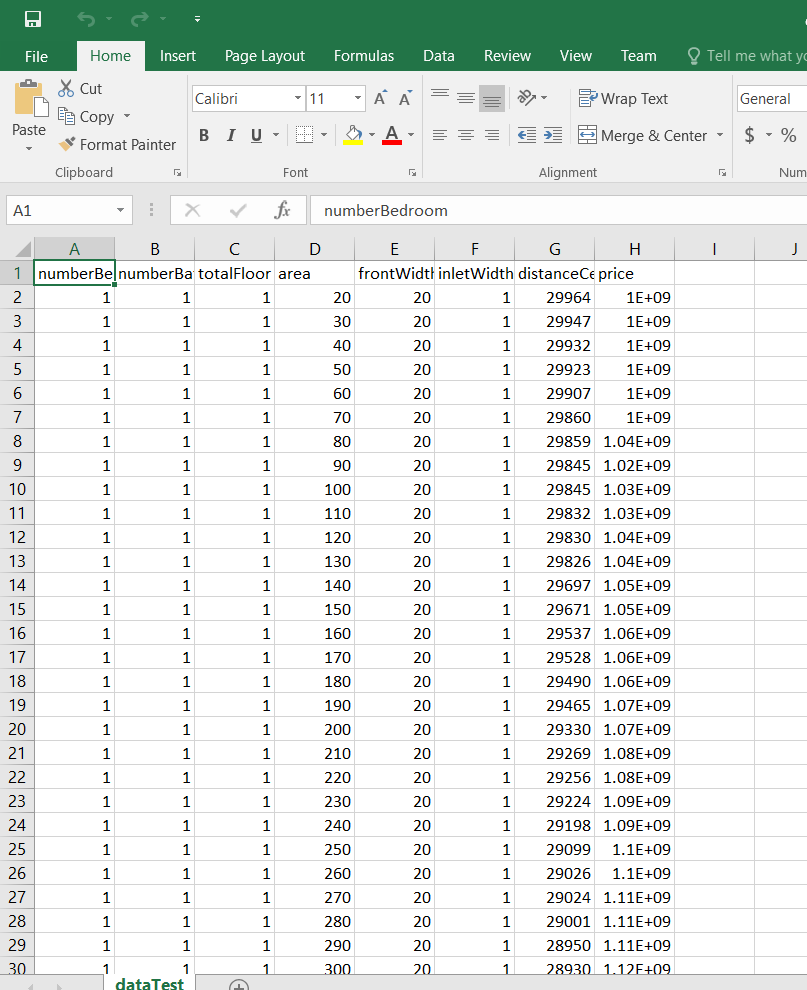


Sau kh lấy được kết quả gợi ý, hệ thống sẽ hiển thị kết quả giá nhà và một số ngôi nhà có giá giao động quanh kết quả đó.

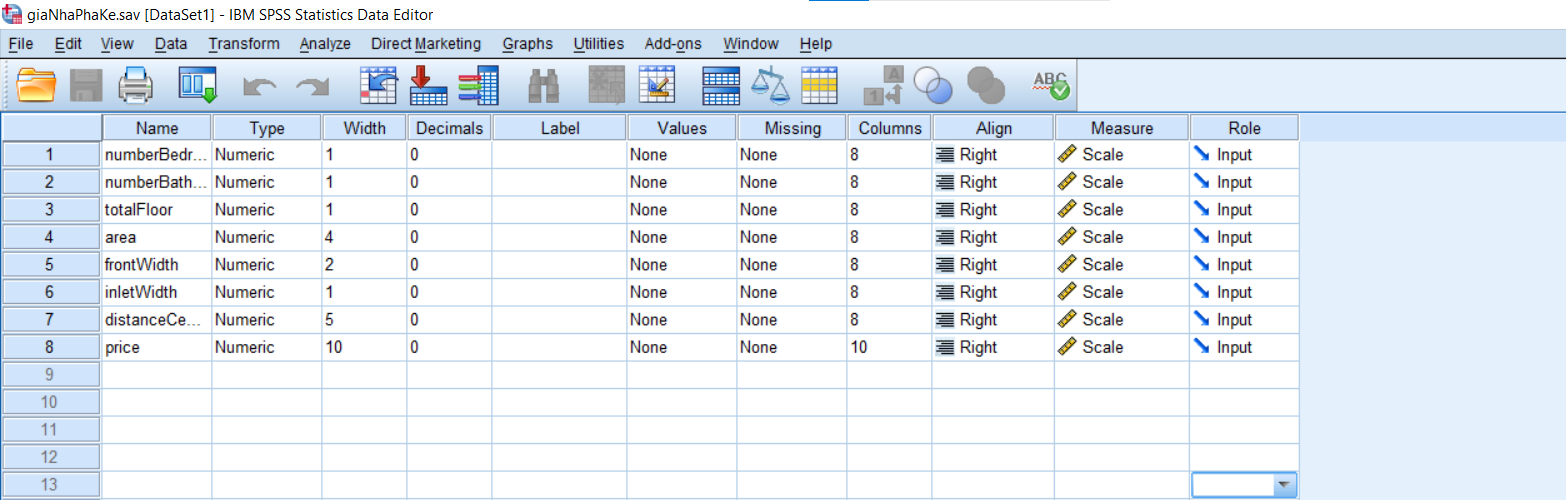


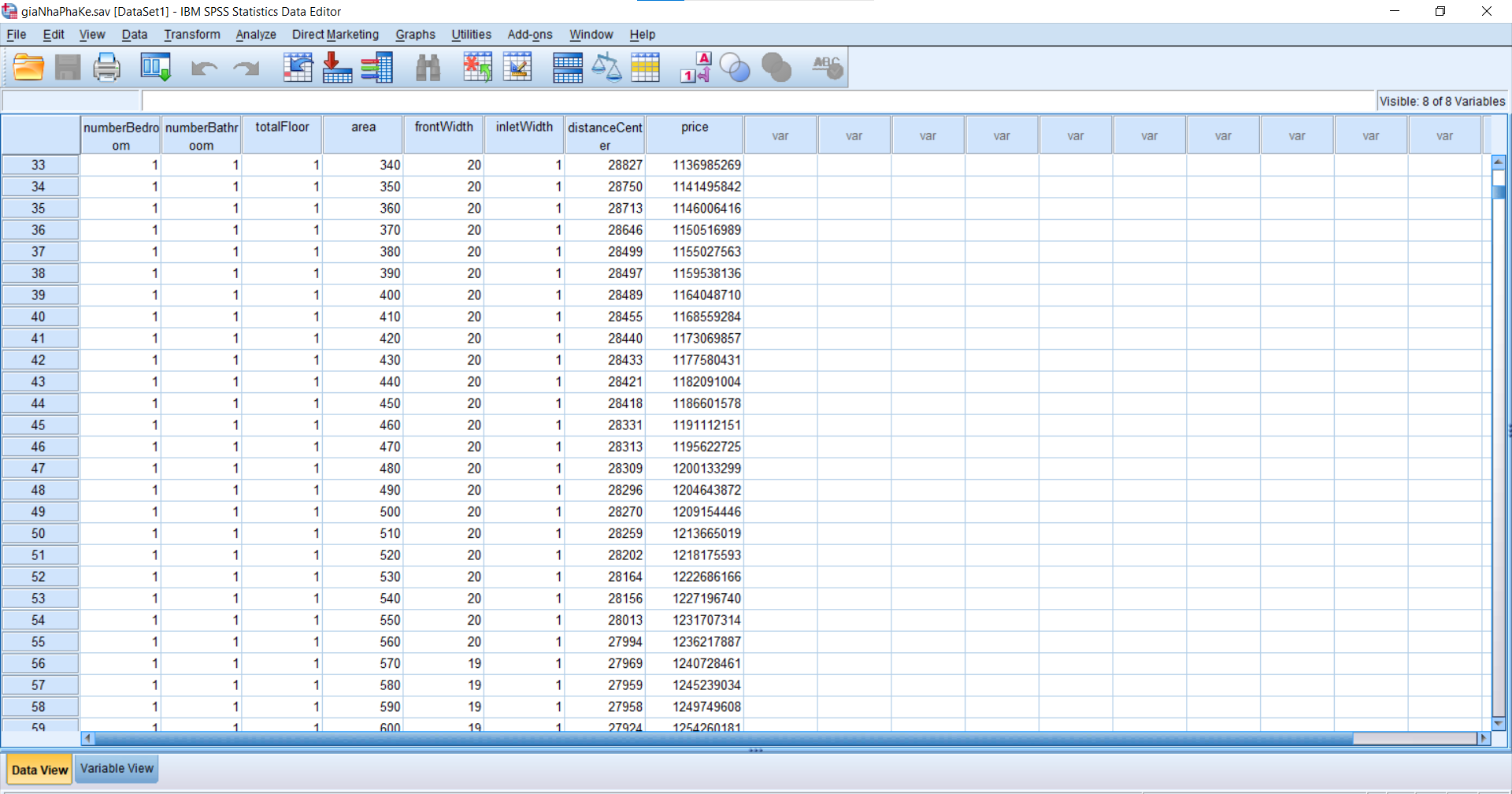
# 3. Phân tích chất lượng dữ liệu

* Sử dụng phần mềm SPSS để phân tích dữ liệu:
* Đã có dữ liệu tại file csv



* Nhập dữ liệu vào phần mềm spss:
* Bằng công cụ hỗ trợ spss đọc dữ liệu dạng số vào bảng

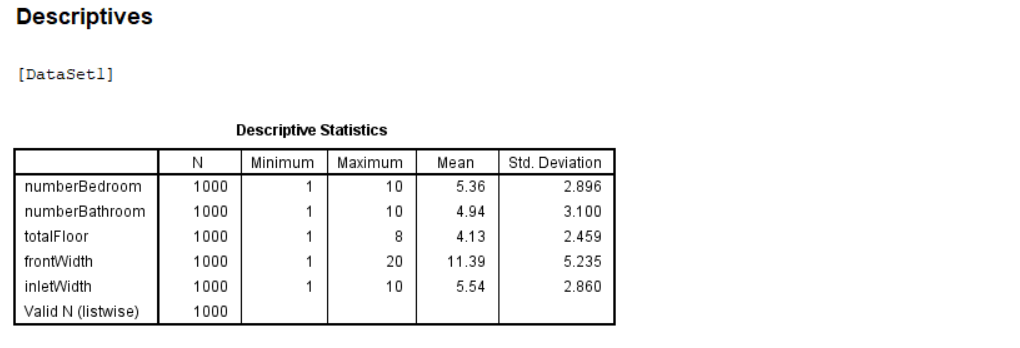


* 
* Để phân tích bộ dữ liệu này e chọn:
* Phân tích tần số dữ liệu:
  + FREQUENCY Analyze – Descriptive Statistics -> Frequencies :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **numberBathroom** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 1 | 214 | 21.4 | 21.4 | 21.4 |
| 2 | 96 | 9.6 | 9.6 | 31.0 |
| 3 | 81 | 8.1 | 8.1 | 39.1 |
| 4 | 85 | 8.5 | 8.5 | 47.6 |
| 5 | 86 | 8.6 | 8.6 | 56.2 |
| 6 | 82 | 8.2 | 8.2 | 64.4 |
| 7 | 87 | 8.7 | 8.7 | 73.1 |
| 8 | 92 | 9.2 | 9.2 | 82.3 |
| 9 | 86 | 8.6 | 8.6 | 90.9 |
| 10 | 91 | 9.1 | 9.1 | 100.0 |
| Total | 1000 | 100.0 | 100.0 |  |

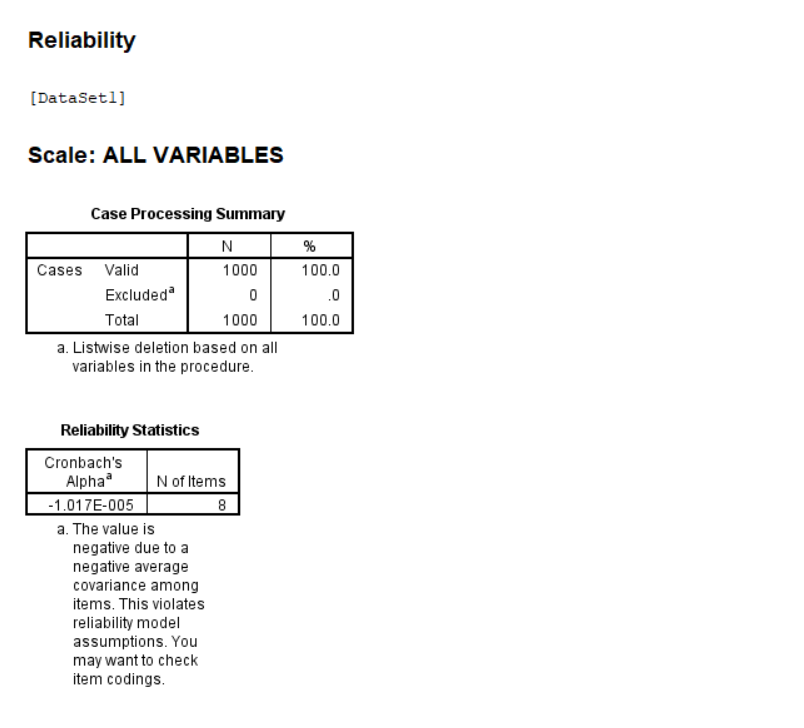
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **totalFloor** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 1 | 235 | 23.5 | 23.5 | 23.5 |
| 2 | 100 | 10.0 | 10.0 | 33.5 |
| 3 | 108 | 10.8 | 10.8 | 44.3 |
| 4 | 99 | 9.9 | 9.9 | 54.2 |
| 5 | 110 | 11.0 | 11.0 | 65.2 |
| 6 | 121 | 12.1 | 12.1 | 77.3 |
| 7 | 113 | 11.3 | 11.3 | 88.6 |
| 8 | 114 | 11.4 | 11.4 | 100.0 |
| Total | 1000 | 100.0 | 100.0 |  |

* Phương sai và độ lệch chuẩn của dữ liệu:
  + DESCRIPTIVE Analyze – Descriptive Statistics -> Descriptives



Giá trị trung bình từng trường dữ liệu và độ lệch chuẩn được tính ở 2 cột cuối giá trị độ lệch chuẩn của bộ dữ liệu này rất lớn chứng tỏ dải dữ liệu rất rộng và trải dài.

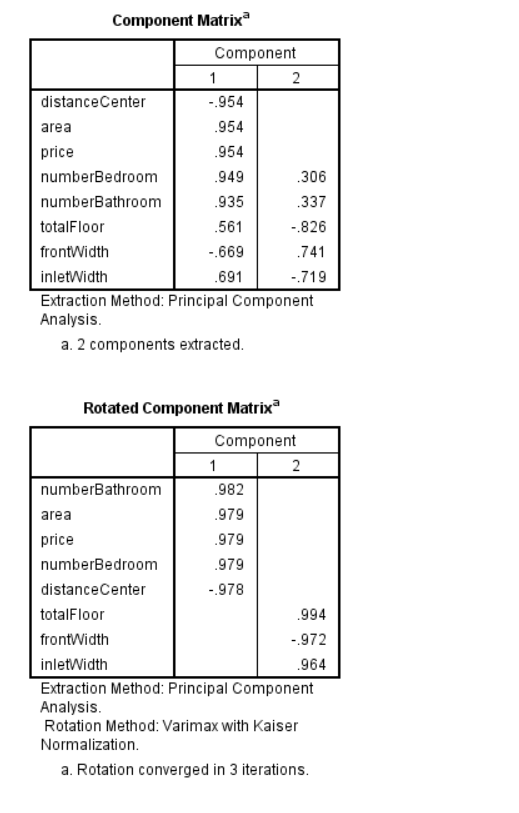
* Phân tích độ tin cậy dữ liệu:
  + Analyze -> Scale -> Reliabilites Analysis



Do bộ dữ liệu này bọn em generate auto từ 1 phần mềm t3 nên độ chính xác, tin cậy dữ liệu :

Đã tính ra âm theo thuật toán và đã báo lỗi dữ liệu và độ chính xác của việc lấy mẫu.

* Phân tích nhân tố khám phá :
  + Analyze -> Dimention Reduction -> Factor:
  + Phân tích này dùng để xét xem các thuộc tính biến không phụ thuộc có liên quan đến nhau hoặc cùng thuộc 1 nhóm



# 4. Trực quan hóa dữ liệu

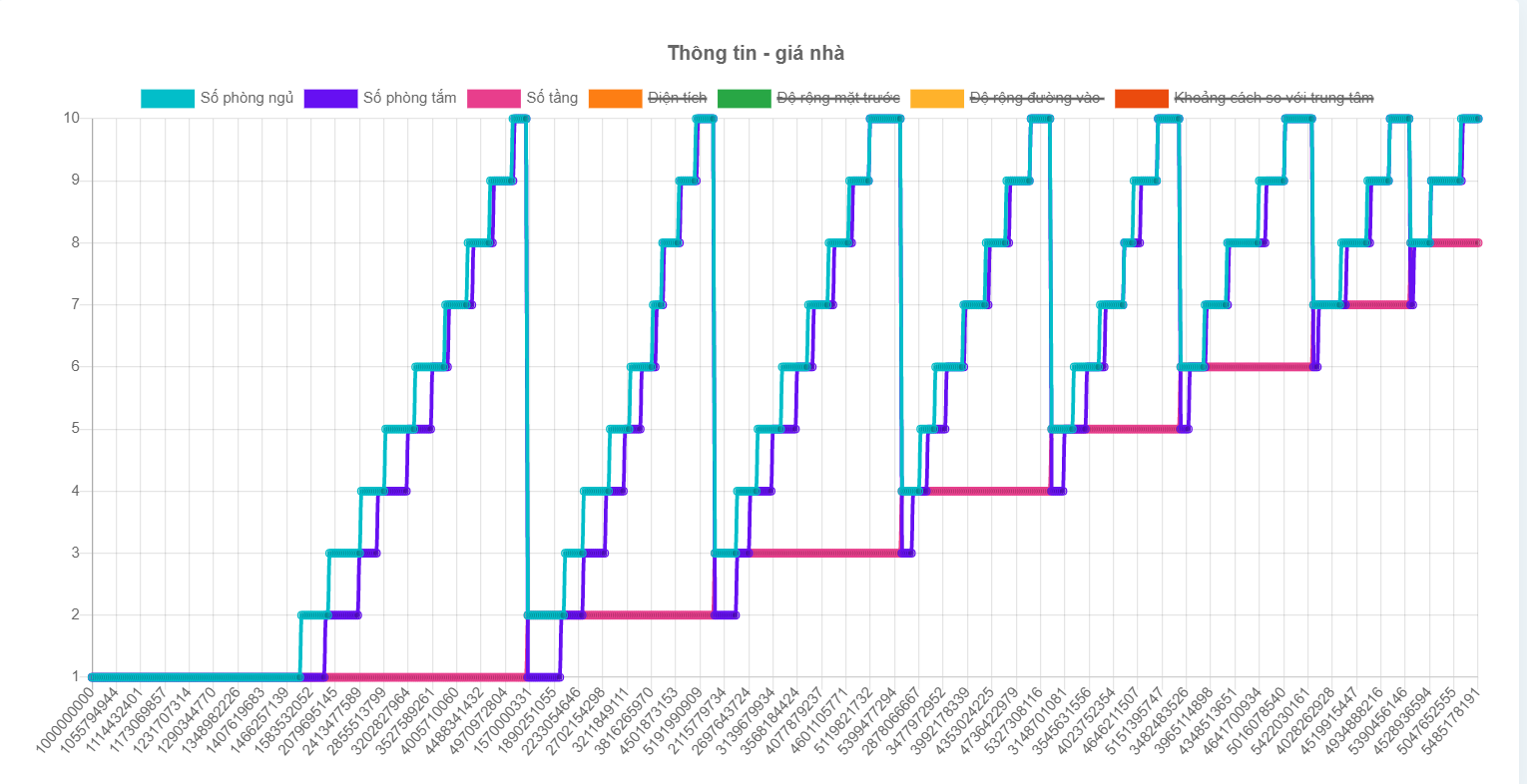
Xây dựng công cụ ng-chart của Angular 7 để trực quan hóa dữ liệu:

- Biểu đồ tổng quan của toàn bộ dữ liệu đối với giá nhà.

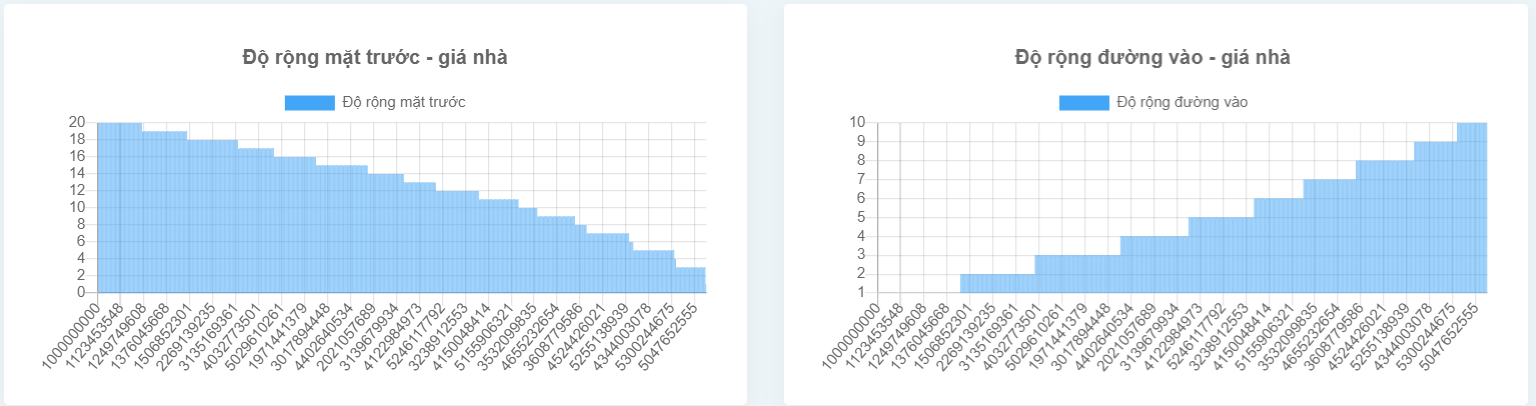
- Biểu đồ chi tiết của từng dữ liệu đối với giá nhà:

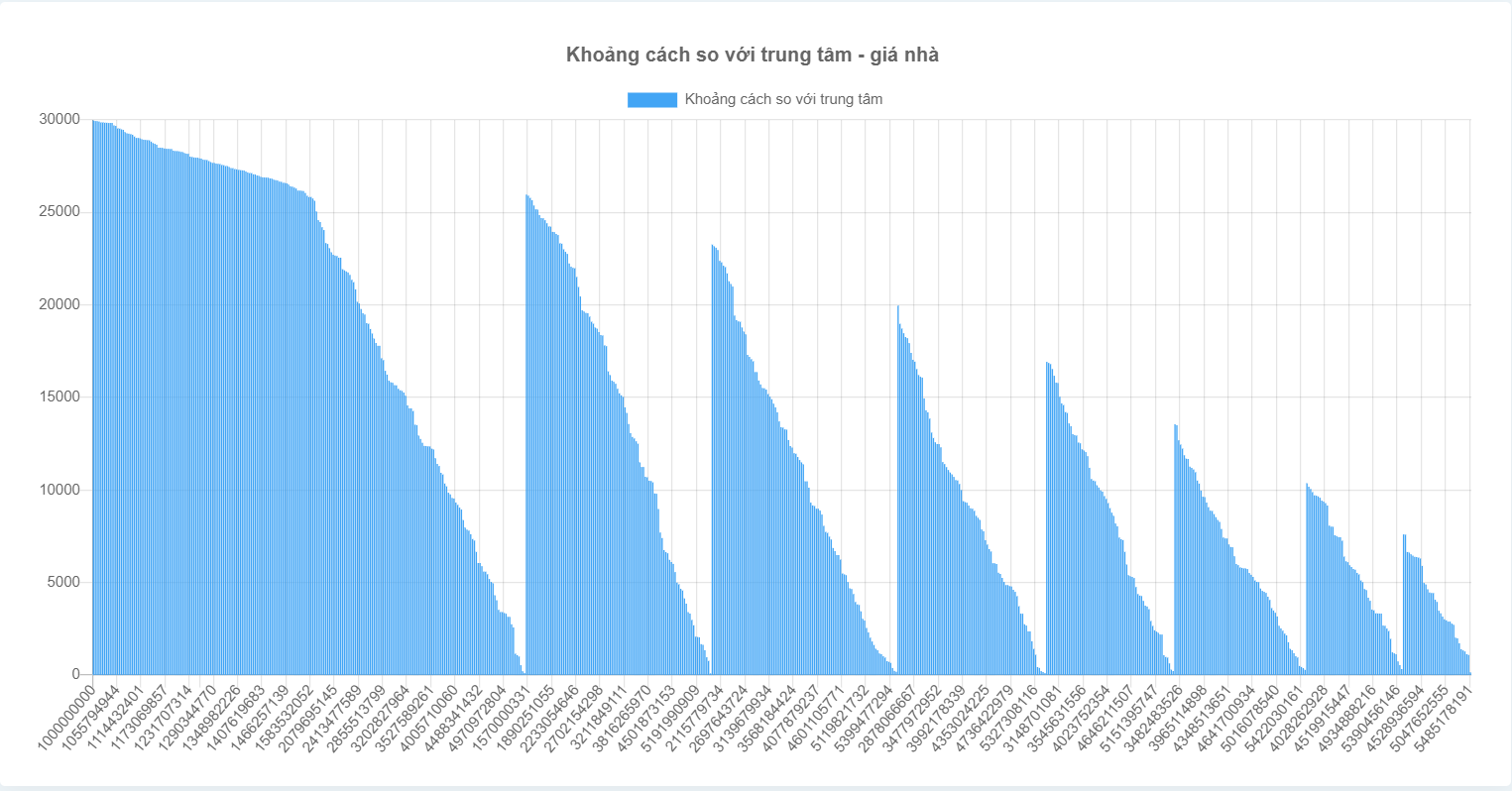


Các biểu đồ:









# 5. Đề xuất giải pháp làm sạch dữ liệu

Đề xuất phương pháp làm sạch theo các tiêu chí:

5.1. Data Consistency: Tính thống nhất của dữ liệu

Các bản ghi không thống nhất thông tin là các bản ghi có số phòng ngủ nhỏ hơn số phòng tắm, số phòng ngủ nhỏ hơn số tầng

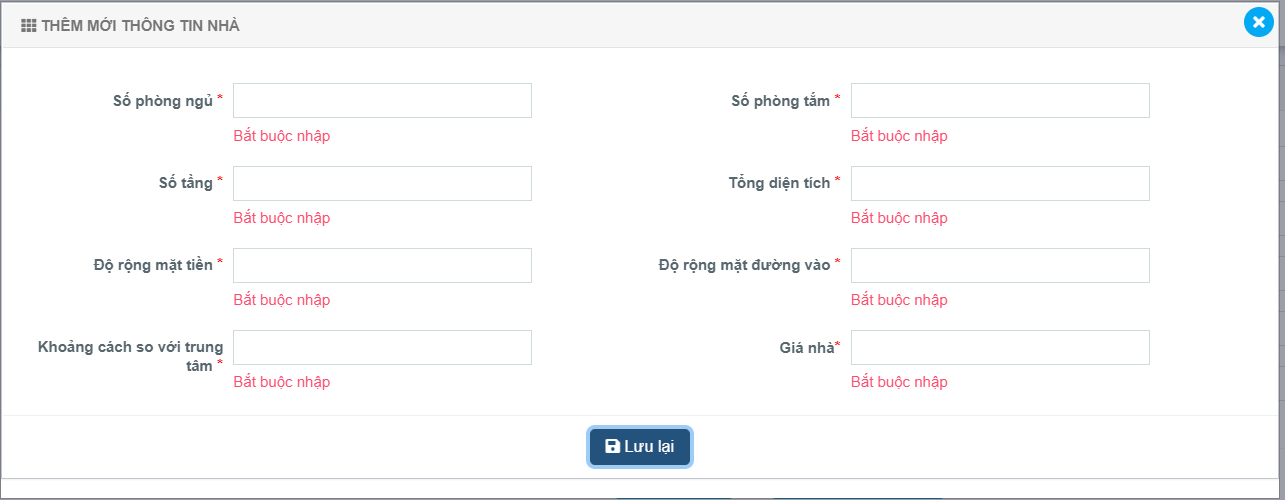
Làm sạch dữ liệu bằng cách xóa các bản ghi có dữ liệu như trên.



5.2. Completenese information: Sự hoàn thiện của dữ liệu

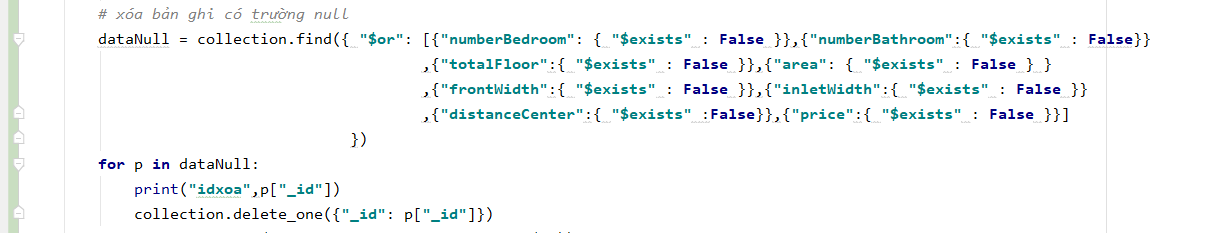
Các bản ghi khi được thêm vào có thể bị thiếu dữ liệu các trường thông tin.

Với trường hợp thêm mới hoặc cập nhật lại trên màn hình, sẽ yêu cầu nhập đầy đủ các thông tin.



Với các bản ghi được import hoặc các lý do khác có thể làm mất mát, thiếu dữ liệu

Thực hiện xóa bỏ các bản ghi đó.



5.3. Data accurercy : Độ chính xác của dữ liệu:

Các dữ liệu được nhập vào phải thỏa mãn các điều kiện nằm trong các khoảng giá trị được quy định. Kiểu dữ liệu là kiểu số

