

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**Khoa Công nghệ Thông tin**



# **TRỰC QUAN HÓA DỮ LIỆU**

## **ĐỒ ÁN CUỐI KỲ**

Hồ Chí Minh - 3/2024

# MỤC LỤC

DANH MỤC ẢNH	i
--------------	---

DANH MỤC BẢNG	ii
---------------	----

<b>1</b>	<b>Giới thiệu</b>	<b>1</b>
1.1	Chủ đề . . . . .	1
1.2	Tập dữ liệu . . . . .	1
1.3	Tiền xử lý dữ liệu . . . . .	1
<b>2</b>	<b>Các thông số thống kê</b>	<b>2</b>
2.1	Thống kê các biến định lượng . . . . .	2
2.1.1	Bảng biểu thống kê . . . . .	2
2.1.2	Các nhận xét . . . . .	2
2.2	Phân tích về các biến định tính . . . . .	4
2.2.1	Bảng biểu thống kê . . . . .	4
2.2.2	Các nhận xét . . . . .	4
<b>3</b>	<b>Phân tích</b>	<b>5</b>
3.1	Dashboard . . . . .	5
3.2	Tổng quan . . . . .	5
3.3	Biểu đồ cột . . . . .	6
3.4	Biểu đồ đường . . . . .	7
3.5	Biểu đồ phân tán . . . . .	8
3.6	Biểu đồ hộp . . . . .	9
3.7	Biểu đồ địa lý . . . . .	10
<b>4</b>	<b>Tương tác với dashboard</b>	<b>11</b>
4.1	Tương tác với biểu đồ cột . . . . .	11
4.2	Tương tác với biểu đồ đường . . . . .	12

4.3	Tương tác với biểu đồ địa lý . . . . .	13
<b>5</b>	<b>Kết luận</b>	<b>13</b>
<b>6</b>	<b>Đánh giá</b>	<b>14</b>

## DANH MỤC HÌNH ẢNH

Hình 2.1	Bảng thống kê các thuộc tính định lượng . . . . .	2
Hình 2.2	Biểu đồ mức độ tác động giữa thuộc tính định lượng . . . . .	3
Hình 2.3	khoảng tin cậy 95% cho giá trị trung bình của mỗi cột: . . . . .	3
Hình 2.4	Bảng thống kê các thuộc tính định tính . . . . .	4
Hình 3.5	Tổng quan về dashboard . . . . .	5
Hình 3.6	Tổng quan về dữ liệu . . . . .	5
Hình 3.7	Biểu đồ cột . . . . .	6
Hình 3.8	Biểu đồ đường . . . . .	7
Hình 3.9	Biểu đồ phân tán . . . . .	8
Hình 3.10	Biểu đồ hộp . . . . .	9
Hình 3.11	Biểu đồ địa lý . . . . .	10
Hình 4.12	Tương tác với biểu đồ cột . . . . .	11
Hình 4.13	Tương tác với biểu đồ đường . . . . .	12
Hình 4.14	Tương tác với biểu đồ địa lý . . . . .	13

# DANH MỤC BẢNG BIỂU

# 1 Giới thiệu

## 1.1 Chủ đề

Hãy phân tích và trực quan hoá các dữ liệu liên quan đến đất nước Việt Nam.

Chủ đề lựa chọn: Phân tích và trực quan số liệu nhà ở tại thủ đô Hà Nội.

## 1.2 Tập dữ liệu

Dữ liệu được lấy từ kho dữ liệu trực tuyến mã nguồn mở **Kaggle**, truy cập bộ dữ liệu tại đây.

Dữ liệu gốc gồm có 12 thuộc tính bao gồm: 6 thuộc tính định tính và 6 thuộc tính định lượng như sau:

- Ngày: Ngày ghi nhận về thông tin căn nhà này.
- Địa chỉ: địa chỉ của căn nhà này tại thủ đô Hà Nội.
- Quận: Quận (hoặc ngang cấp quận) mà căn nhà này tọa lạc.
- Huyện: Huyện (hoặc ngang cấp Huyện) mà căn nhà này tọa lạc.
- Loại hình nhà ở: Thông tin về loại hình mà căn nhà này thuộc về.
- Giấy tờ pháp lý: Thông tin về giấy tờ hợp pháp của căn nhà này.
- Số tầng: Số tầng của căn nhà này.
- Số phòng ngủ: Số phòng ngủ của căn nhà này.
- Diện tích: thông tin về diện tích của căn nhà.
- Dài: thông tin về chiều dài của căn nhà.
- Rộng: thông tin về chiều rộng của căn nhà.
- Giá: Thông tin về giá tiền trên 1  $m^2$  của căn nhà.

## 1.3 Tiền xử lý dữ liệu

**Loại bỏ chữ:** Ở trong các thuộc tính định lượng như chiều dài rộng, giá, và diện tích. Việc loại bỏ các đơn vị đo lường (triệu, đồng, m,  $m^2$ ) nhằm dễ dàng hơn trong việc trực quan hoá trên nền tảng tableau cũng như tính toán về sau.

**Xử lý trùng lặp:** Trong tập dữ liệu gốc có tới hơn 80.000 dòng dữ liệu nên việc trùng lặp là không thể tránh khỏi, chính vì vậy, việc xử lý trùng lặp cũng khiến cho việc tính toán trở nên công bằng và thực tế hơn.

**Loại bỏ cột không cần thiết:** Có 1 cột "Unnamed: 0" không có giá trị trong dữ liệu cũng

như bài phân tích và trực quan này nên cần được loại bỏ để dữ liệu sạch hơn.

**Loại bỏ giá trị rỗng:** Các giá trị bị thiếu hay không được cung cấp sẽ bị loại bỏ để thuận tiện hơn trong quá trình trực quan dữ liệu trở nên mượt mà hơn.

**Loại bỏ nhiễu:** Dùng phương pháp 1.5 IQR để loại bỏ các giá trị nhiễu để tập trung vào các giá trị phổ biến làm các biểu đồ trực quan dễ quan sát cũng như cung cấp được nhiều thông tin hơn.

**Loại bỏ dữ liệu sai quy định pháp luật:** Các dữ liệu nhà ngỗ, hẻm có số tầng lớn hơn 4; nhà phố liền kề có số tầng lớn hơn 6; nhà biệt thự có số tầng lớn hơn 3; nhà mặt phố mặt tiền có số tầng lớn hơn 8 sẽ bị loại bỏ khỏi bộ dữ liệu để phù hợp với quy định của nhà nước.

## 2 Các thông số thống kê

### 2.1 Thống kê các biến định lượng

#### 2.1.1 Bảng biểu thống kê

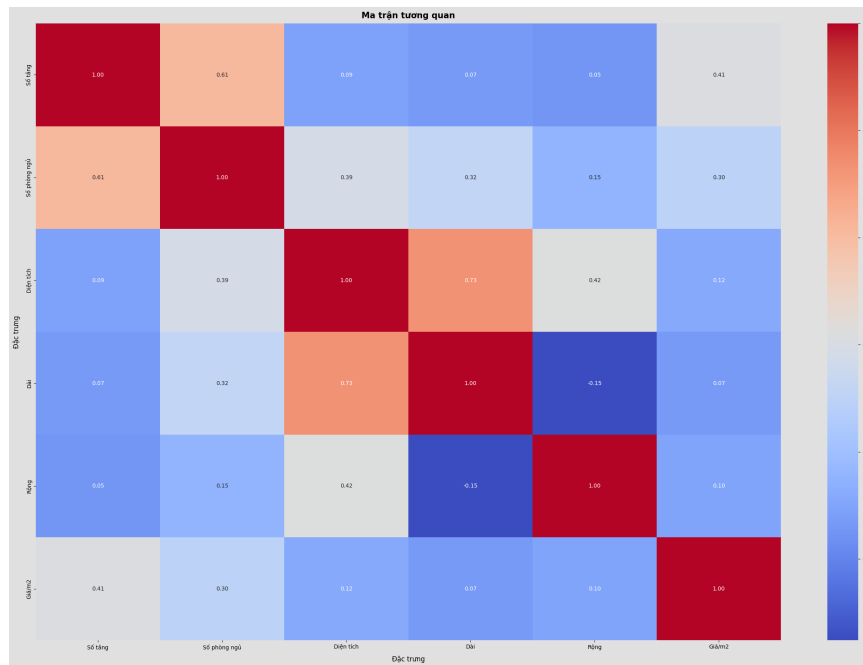
	count	mean	std	min	25%	50%	75%	max
Số tầng	5751.0	3.867849	0.954401	1.0	4.00	4.00	4.00	8.00
Số phòng ngủ	5751.0	3.549470	1.203883	1.0	3.00	4.00	4.00	11.00
Diện tích	5751.0	41.890871	12.927284	4.0	33.00	40.00	50.00	86.00
Dài	5751.0	10.011707	2.682718	3.0	8.00	10.00	12.00	17.80
Rộng	5751.0	4.115239	0.888052	2.0	3.50	4.00	4.80	7.00
Giá/m <sup>2</sup>	5751.0	86.179467	32.550020	0.0	64.52	81.25	103.03	182.69

**Hình 2.1** Bảng thống kê các thuộc tính định lượng

#### 2.1.2 Các nhận xét

Nhận xét về thống kê chung:

- Thông số **Số tầng**: Biểu đồ này cho thấy số tầng của các tòa nhà được khảo sát. Phần lớn các tòa nhà có từ 4 đến 6 tầng, với số lượng tòa nhà 5 tầng là nhiều nhất. Điều này có thể cho thấy một xu hướng chung trong thiết kế xây dựng tại khu vực này, ưu tiên các tòa nhà vừa và nhỏ.
- Thông số **diện tích**: Biểu đồ này thể hiện sự phân bố diện tích của các căn nhà hoặc đất được khảo sát. Dữ liệu tập trung chủ yếu trong khoảng từ 20 đến 60m<sup>2</sup>, với số lượng lớn nhất là khoảng 40m<sup>2</sup>. Điều này có thể cho thấy rằng đa số các căn nhà hoặc mảnh đất ở khu vực khảo sát có diện tích vừa phải, phù hợp với nhu cầu sinh sống của hầu hết các hộ gia đình.
- Thông số **Giá/m<sup>2</sup>**: Biểu đồ này hiển thị giá tiền trên mỗi mét vuông, một thông số quan trọng để đánh giá giá trị bất động sản. Sự phân bố tập trung chủ yếu ở khoảng 50 đến 125 triệu đồng/m<sup>2</sup>, với điểm cao nhất là khoảng 75 triệu đồng/m<sup>2</sup>. Điều này cho thấy giá cả bất động sản trong khu vực này khá đa dạng nhưng phần lớn vẫn ở mức vừa phải.



**Hình 2.2** Biểu đồ mức độ tác động giữa thuộc tính định lượng

khoảng tin cậy 95% cho giá trị trung bình của mỗi cột:		
	Giá trị chặn dưới	Giá trị chặn trên
Số tầng	3.069950	4.665748
Số phòng ngủ	2.542998	4.555941
Diện tích	31.083392	52.698351
Dài	7.768898	12.254515
Rộng	3.372809	4.857669
Giá/m2	58.966969	113.391965

**Hình 2.3** khoảng tin cậy 95% cho giá trị trung bình của mỗi cột:

Nhận xét về tác động giữa các thuộc tính:

Tương quan mạnh:

- Có tương quan cao giữa "Diện tích" và "Dài" với hệ số là 0.73. Điều này có nghĩa là chiều dài của bất động sản thường ảnh hưởng lớn đến diện tích tổng thể, điều này là dễ hiểu vì diện tích thường được tính bằng chiều dài nhân với chiều rộng.
- "Số tầng" và "Số phòng ngủ" cũng có mối tương quan khá cao (0.61), cho thấy rằng số lượng tầng trong một căn nhà thường có liên quan đến số lượng phòng ngủ.

Tương quan trung bình:

- "Số phòng ngủ" và "Diện tích" có hệ số tương quan là 0.39, cho thấy một mối quan hệ tích cực nhưng không quá mạnh. Điều này có thể do các căn nhà lớn hơn có nhiều phòng ngủ hơn nhưng không phải lúc nào cũng vậy.



- "Dài" và "Đặc trưng" có hệ số tương quan là 0.42, cho thấy một mối quan hệ vừa phải giữa đặc trưng này và kích thước chiều dài của bất động sản.

Tương quan thấp hoặc không đáng kể:

- Nhiều cặp biến như "Số tầng" với "Diện tích" hoặc "Dài" với "Số phòng ngủ" cho thấy hệ số tương quan thấp (dưới 0.15), cho thấy không có mối quan hệ rõ ràng hoặc mạnh mẽ giữa các biến này.
- "Giá/m<sup>2</sup>" có tương quan thấp với hầu hết các đặc trưng khác, cho thấy giá trên mỗi mét vuông không phụ thuộc nhiều vào kích thước hoặc cấu trúc của bất động sản.

## 2.2 Phân tích về các biến định tính

### 2.2.1 Bảng biểu thống kê

	num_values	value_ratios
Ngày	79	{'2020-08-04': 2.7, '2020-07-24': 2.6, '2020-0...
Địa chỉ	2512	{'Đường Khương Trung, Phường Khương Trung, Quậ...
Quận	22	{'Hà Đông': 19.5, 'Đống Đa': 12.5, 'Thanh Xuân...
Huyện	202	{'Phường Yên Nghĩa': 3.9, 'Phường Khương Trung...
Loại hình nhà ở	4	{'Nhà ngỗ, hẻm': 66.5, 'Nhà mặt phố, mặt tiền'...
Giấy tờ pháp lý	4	{'Đã có sổ': 97.8, 'nan': 0.8, 'Đang chờ sổ': ...
Location	22	{'Hà Nội, Quận Hà Đông': 19.5, 'Hà Nội, Quận Đ...

Hình 2.4 Bảng thống kê các thuộc tính định tính

### 2.2.2 Các nhận xét

Nhận xét về thống kê chung:

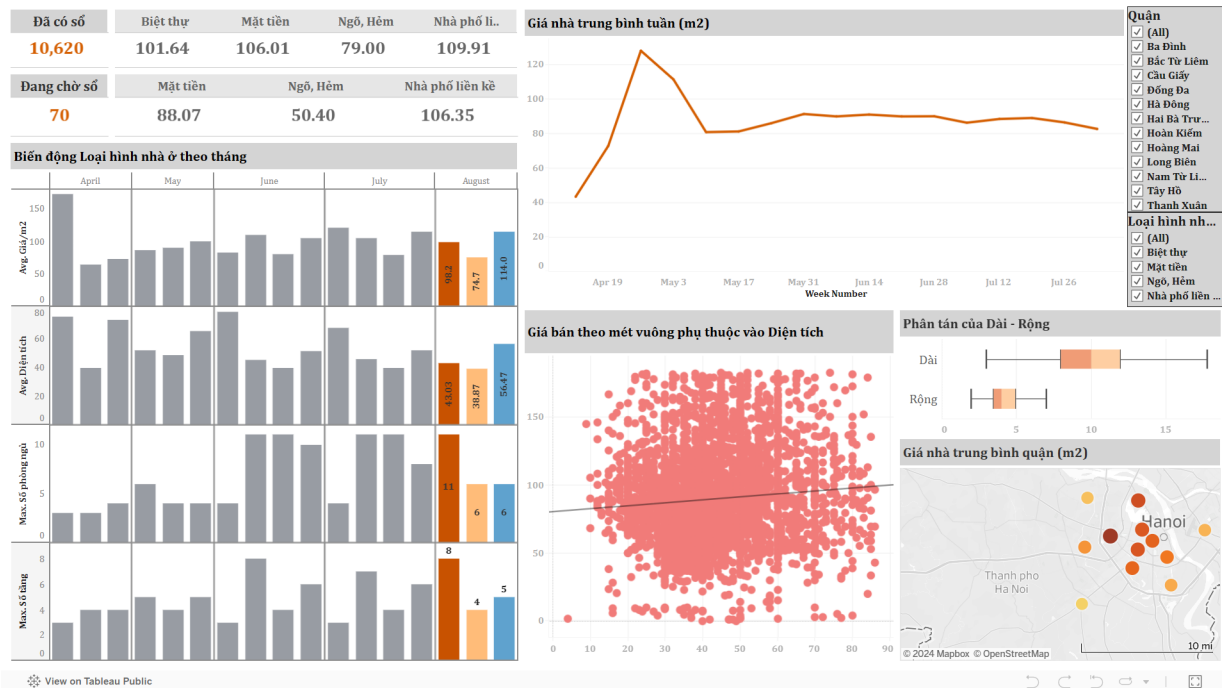
- Thông số **Loại hình nhà ở**: Biểu đồ này thể hiện sự phân bố số lượng theo loại hình nhà ở. Phần lớn các tài sản là "Nhà riêng lẻ", tiếp theo là "Nhà liền kề" với số lượng đáng kể thấp hơn, và cuối cùng là "Nhà biệt thự", "Nhà tập thể" có số lượng ít nhất. Sự chênh lệch lớn giữa các loại hình nhà ở cho thấy phong cách và lựa chọn ở của người dân trong khu vực. Điều này có thể phản ánh nhu cầu và khả năng kinh tế của cư dân, cũng như xu hướng phát triển đô thị tại các khu vực đó. Những thông tin này rất quan trọng trong việc phân tích thị trường bất động sản và lập kế hoạch đô thị.
- Thông số **Giấy tờ pháp lý**: Biểu đồ này thể hiện số lượng giấy tờ pháp lý của các nhà, đất được phân theo loại. Sự chênh lệch lớn giữa những tài sản có giấy tờ đầy đủ và những tài sản không có hoặc chỉ có sổ tạm, cho thấy mức độ pháp lý của bất động sản tại khu vực này. Tình trạng này có thể ảnh hưởng lớn đến giá trị và khả năng giao dịch của bất động sản.

- Thông số **Quận**: Biểu đồ này cho thấy sự phân bố số lượng theo các quận trong thành phố. Quận 1 và Quận 7 nổi bật với số lượng cao nhất, cho thấy sự tập trung dân cư hoặc hoạt động thương mại đáng kể trong các khu vực này. Điều này có thể phản ánh mức độ phát triển và sự ưu tiên trong đầu tư cơ sở hạ tầng và dịch vụ.

## 3 Phân tích

### 3.1 Dashboard

Có thể truy cập đầy đủ dashboard tại đây



Hình 3.5 Tổng quan về dashboard

### 3.2 Tổng quan

Đã có số	Biệt thự	Mặt tiền	Ngõ, Hẻm	Nhà phố li..
10,620	101.64	106.01	79.00	109.91
Đang chờ số	Mặt tiền	Ngõ, Hẻm	Nhà phố liền kề	
70	88.07	50.40	106.35	

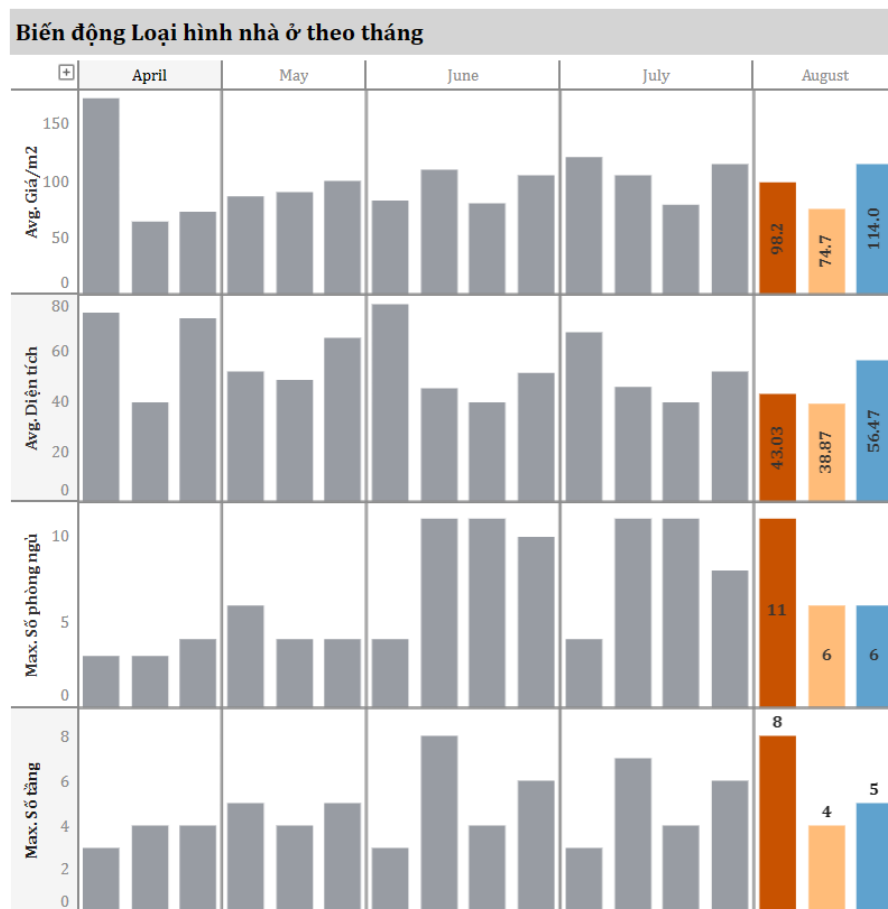
Hình 3.6 Tổng quan về dữ liệu

Bảng thống kê thể hiện các số liệu tổng quan liên quan đến tập dữ liệu. Mục đích cung cấp một vài thông tin sơ bộ về dữ liệu các thông tin bao gồm: số lượng nhà đã có sổ, số lượng nhà chưa

có sổ, trung bình giá nhà ở theo từng loại hình nhà ở (mặt tiền, ngõ hẻm, nhà phố liền kề). Ở đây ta lưu tâm đến loại hình nhà ở vì loại hình nhà ở là yếu tố ảnh hưởng nhiều nhất đến giá nhà. Ngoài ra phần này còn giúp so sánh giá nhà của nhà đã có sổ và đang chờ sổ. Bên cạnh đó bảng thống kê còn giúp ta so sánh được các giá nhà trung bình của từng loại hình nhà với nhau giúp có cái nhìn tổng thể về giá nhà.

### 3.3 Biểu đồ cột

Ở phần này ta sẽ so sánh và đánh giá sự khác nhau về giá nhà trung bình, diện tích trung bình, số lượng phòng ngủ, số tầng của từng loại hình nhà ở theo từng tháng. Biểu đồ được sử dụng là biểu đồ cột.



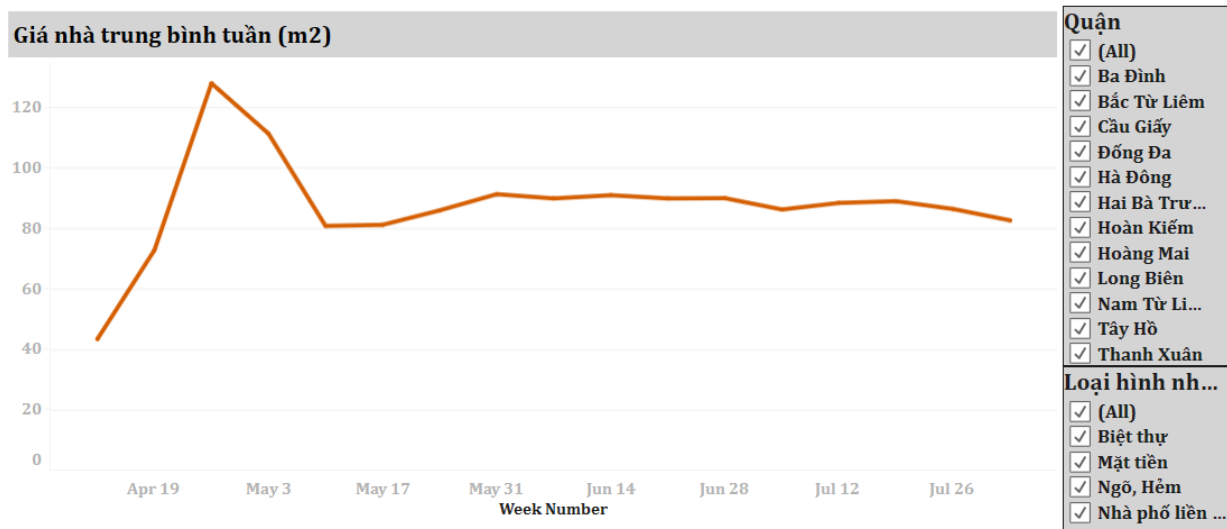
**Hình 3.7** Biểu đồ cột

**Lý do sử dụng:** Biểu đồ cung cấp đầy đủ các thông tin cần trình bày và giúp ta có thể so sánh và đánh giá sự tăng giảm trong từng tháng đối với từng yếu tố (giá nhà, diện tích, số phòng và số lầu) đối với từng loại nhà ở. Ngoài ra việc highlight tháng gần nhất giúp ta xác định được các số liệu sát với thời gian thực có cái nhìn chính xác hơn về số liệu.

**Nhận xét:** Nhìn chung ta có thể thấy được giá trung bình và diện tích trung bình ít và số tầng sẽ ít bị biến động theo từng tháng. Và biến thiên của nó cũng không quá lớn. Bên cạnh đó ta có thể thấy được sự chênh lệch về giá nhà, diện tích, số phòng, số tầng của từng loại hình nhà ở. Nó phù hợp đối với nhu cầu của người dùng.

### 3.4 Biểu đồ đường

Ở phần này chúng ta sẽ thể hiện xu hướng giá trung bình trên 1 mét vuông theo mỗi tuần khác nhau trong một khoảng thời gian nhất định. Biểu đồ được sử dụng là biểu đồ đường.



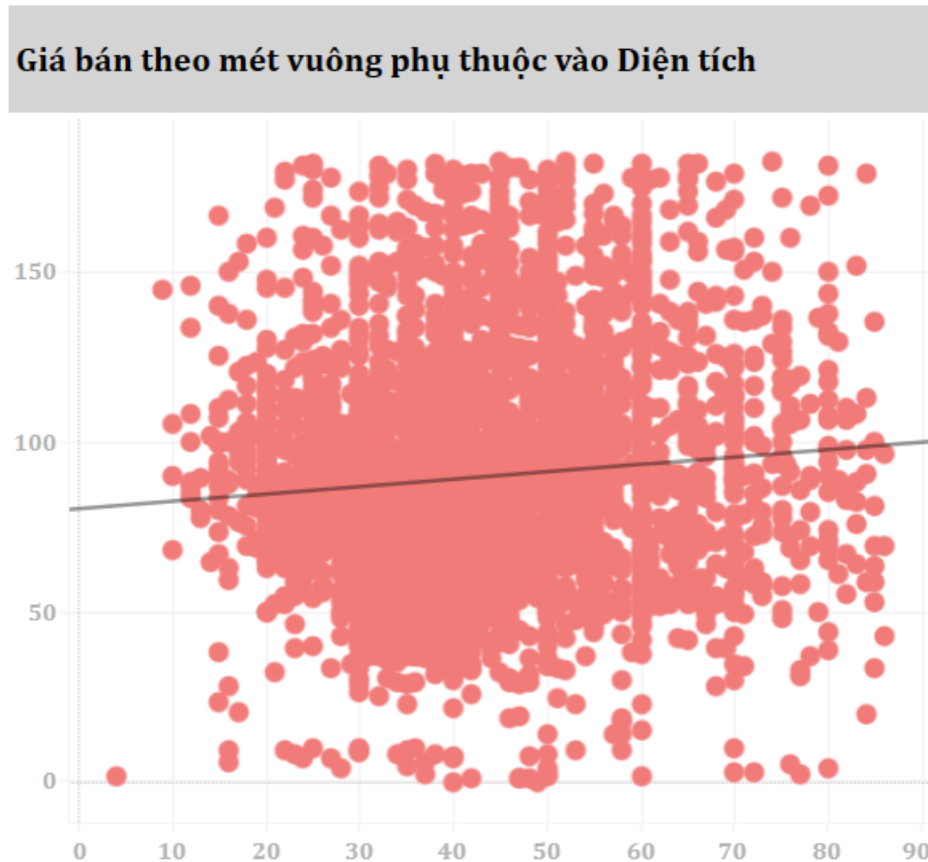
Hình 3.8 Biểu đồ đường

**Lý do sử dụng:** Biểu đồ đường được sử dụng để minh họa xu hướng giá trung bình trên mỗi mét vuông của các căn nhà theo từng tuần. Bằng cách tương tác với bảng lọc, thông qua biểu đồ, chúng ta có thể quan sát xu hướng giá trong các quận và so sánh giữa chúng theo thời gian. Thông qua việc này, chúng ta có thể phân tích các yếu tố như tình hình xã hội, an ninh, và cách chúng ảnh hưởng đến giá nhà. Ngoài ra, chúng ta cũng có thể dự đoán xu hướng giá nhà ở trong tương lai dựa trên các dấu hiệu từ biểu đồ.

**Nhận xét::** Tùy thuộc vào từng quận, thời gian thay đổi sẽ khác nhau, do đó, chúng ta sẽ tiến hành đánh giá tổng quan cho toàn bộ các quận. Phổ quát, giá trung bình thường duy trì ổn định trong dài hạn và đạt đỉnh vào tháng 5. Ban đầu, giá trị trung bình có xu hướng tăng lên vào cuối vài tuần của tháng 4, đạt đỉnh vào tuần đầu tiên của tháng 5 với mức giá trị 106,48 triệu đồng/mét vuông. Tuy nhiên, giá cũng có xu hướng giảm sau khi đạt đỉnh, và cho đến đầu tháng 6 mới bắt đầu tăng nhẹ lên và duy trì ổn định cho đến tháng 8. Tổng kết lại, biểu đồ này cung cấp nhiều thông tin hữu ích về khoảng thời gian mà giá nhà đất dao động mạnh do ảnh hưởng của các yếu tố bên ngoài. Điều này giúp người mua có cái nhìn tổng quan và đưa ra quyết định hợp lý.

### 3.5 Biểu đồ phân tán

Phần này ta sẽ thể hiện sự tương quan giữa diện tích và giá nhà trung bình trên 1 mét vuông. Và có 1 đường thẳng giúp dự đoán giá nhà thuộc thủ đô Hà Nội trong tương lai.

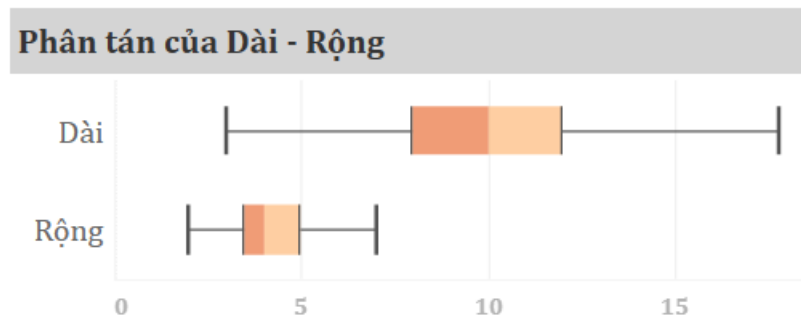


**Hình 3.9** Biểu đồ phân tán

**Lý do sử dụng:** Biểu đồ cho ta thấy được giá nhà trung bình theo mét vuông ứng với từng diện tích nhà. Ngoài ra giúp ta dễ dàng thấy được sự tương quan, phân bố giữa diện tích và giá nhà.

**Nhận xét:** Nhìn chung ta có thể thấy diện tích nhà ít ảnh hưởng đến giá nhà khi ta xét toàn bộ Hà Nội nguyên nhân là do ở vùng ngoại ô thì giá nhà sẽ thấp hơn so với trong nội thành. Ngoài ra loại hình nhà ở cũng là một nguyên nhân quan trọng ảnh hưởng đến giá và làm giảm bớt sự tương quan giữa diện tích và giá nhà.

### 3.6 Biểu đồ hộp

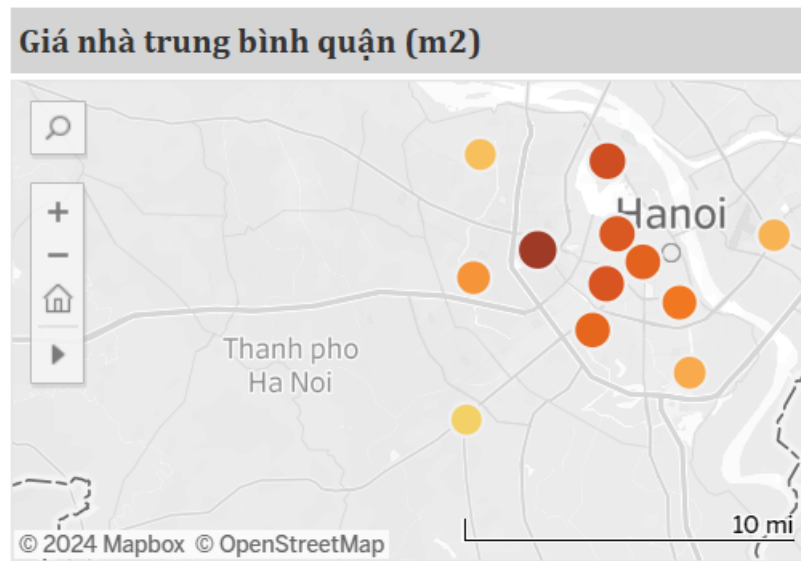


**Hình 3.10** Biểu đồ hộp

**Lý do sử dụng:** Biểu đồ hộp được sử dụng để minh họa các đại lượng của hai miền giá trị: **Dài** (chiều dài của căn nhà) và **Rộng** (chiều rộng của căn nhà). Thông qua biểu đồ này, chúng ta có thể đánh giá các giá trị của từng miền và phân tích phân phối của chúng để xem xét sự lệch lạc và tỷ lệ nhiều của miền giá trị tương ứng.

**Nhận xét:** Nói chung, biểu đồ hộp của từng phạm vi phân phối dữ liệu cho thấy một sự đồng đều khá đáng chú ý, tuy nhiên vẫn có sự xuất hiện của nhiều giá trị nhiễu. Đối với phạm vi Dài, dữ liệu chủ yếu tập trung đều ở hai phía của giá trị trung bình, biểu hiện xu hướng phân bố chuẩn. Trong khi đó, đối với phạm vi Rộng, dữ liệu phân bố đều đặn hơn nhưng có sự chênh lệch nhẹ về phía trên của giá trị trung bình, có thể là theo phân phối chuẩn hoặc xu hướng âm. Phân phối chủ yếu của dữ liệu Rộng tập trung ở phía trên của biểu đồ hộp. Tóm lại, biểu đồ hộp cung cấp một cái nhìn tổng quan về sự phổ biến của chiều dài và chiều rộng của căn nhà, cùng với một số căn nhà có kích thước đặc biệt.

### 3.7 Biểu đồ địa lý



**Hình 3.11** Biểu đồ địa lý

**Lý do sử dụng:** Biểu đồ địa lý cho ta được một thông tin mà các biểu đồ khác không thể cho được chính là vị trí địa lý của các điểm dữ liệu. Theo như ta đã được biết về vị trí nhà ở có thể ảnh hưởng rất lớn đến giá trị của căn nhà đó. Biểu đồ địa lý cho thông tin tương quan giữa vị trí căn nhà đang tọa lạc và giá trị của căn nhà đó.

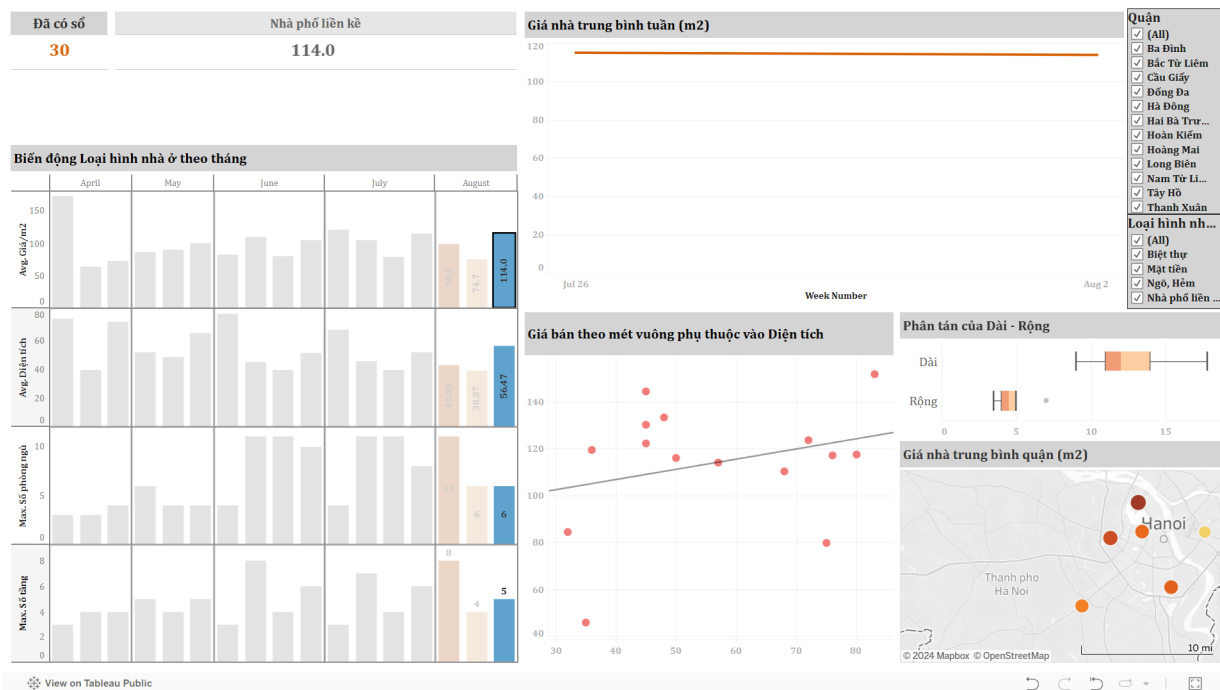
Ở biểu đồ địa lý này, có 2 ánh xạ được sử dụng:

- **Ánh xạ vị trí:** Như đã đề cập phía trên, ta có một ánh xạ từ các điểm dữ liệu với giá trị là các quận của Hà Nội. Qua một số bước tiền xử lý, các tên quận này được chuyển đổi qua 2 trục tọa độ tương ứng và được biểu diễn lên biểu đồ như hình.
- **Ánh xạ về màu sắc:** Với biểu đồ địa lý, chỉ mỗi ánh xạ về vị trí là không đủ để diễn tả được hết ý nghĩa mà biểu đồ này muốn truyền đạt. Do đó một ánh xạ khác đã được thêm vào, ánh xạ từ độ lớn của dữ liệu qua màu sắc. Nếu không có ánh xạ này, các điểm dữ liệu đơn giản là cùng một màu sắc và ta không thể kết luận được điểm dữ liệu nào có độ lớn hơn các điểm dữ liệu khác. Cụ thể trong biểu đồ địa lý ta sử dụng phía trên, nhìn vào màu sắc đậm nhạt ta có thể đưa ra so sánh giá nhà trung bình giữa các quận một cách nhanh chóng.

**Nhận xét:** Biểu đồ cho ta thấy giá trị của căn nhà khi nằm ở trung tâm thủ đô Hà Nội có xu hướng cao hơn ở vùng ven thủ đô. Đây là một điều dễ hiểu khi đa số dân cư, các công ty, các trung tâm hành chính đều tập trung về hướng trung tâm thủ đô và do đó kéo giá nhà khu vực này tăng cao. Khu vực ven vùng trung tâm có mật độ dân cư thưa thớt hơn, do đó nhu cầu về nhà ở không cao nên giá thành sẽ thấp hơn nhiều so với khu vực trung tâm.

## 4 Tương tác với dashboard

### 4.1 Tương tác với biểu đồ cột



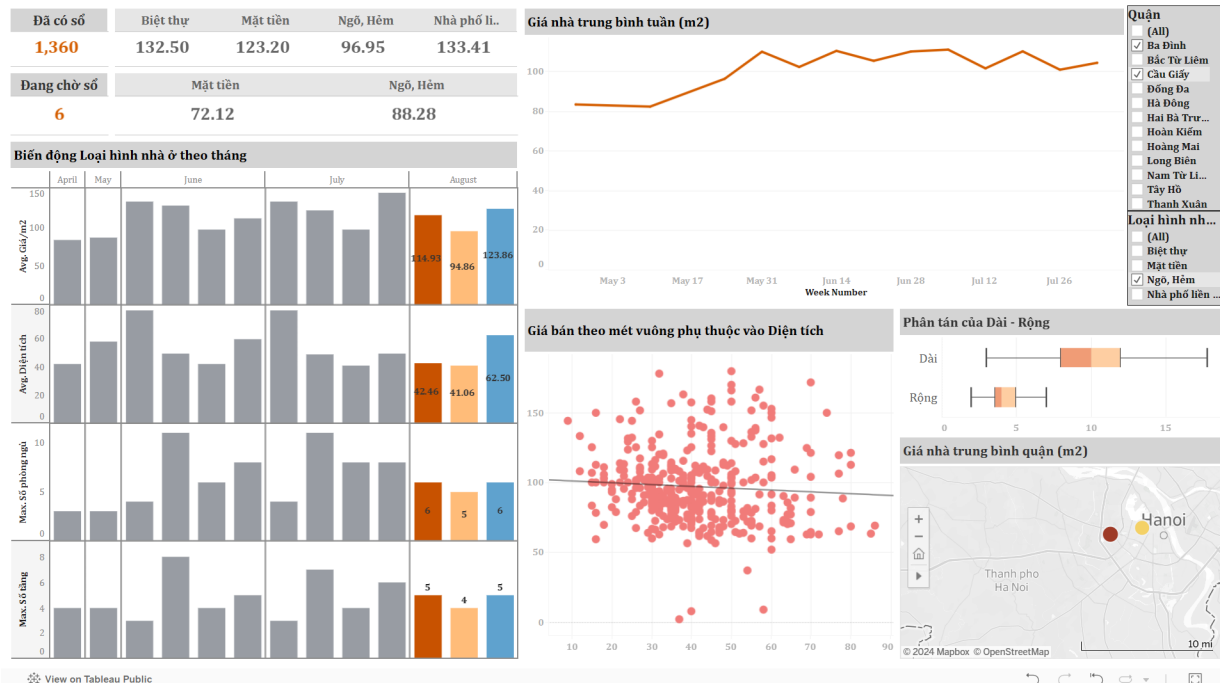
Hình 4.12 Tương tác với biểu đồ cột

Ở biểu đồ cột ta có thể chọn ra loại hình nhà ở trong một tháng cụ thể hoặc tất cả loại hình nhà ở trong tháng đó. Các thông số ở các biểu đồ khác như phần thống kê, biểu đồ đường, biểu đồ phân tán sẽ được tự động thay đổi theo. Như trong hình bên dưới, khi chọn loại hình biệt thự được rao bán vào tháng 8, ta thấy được một số thông tin cụ thể như số lượng, vị trí ở biểu đồ địa lý, biến động giá bán theo thời gian,...

**Nhận xét:** Việc sử dụng hình thức tương tác này cho người dùng có thể sử dụng dashboard một cách linh hoạt hơn. Đôi khi những thông tin được hiển thị trên dashboard không đủ cho người dùng có một cái nhìn sâu sắc về dữ liệu. Lấy ví dụ như việc người dùng chỉ muốn xem giá nhà trong tháng 8 thay vì toàn bộ 2 năm như dashboard mặc định hiển thị, nếu không có sự tương tác này, người dùng không thể có được thông tin mình cần. Do đó đây là một cách hay và hợp lý để giải quyết vấn đề trên.



## 4.2 Tương tác với biểu đồ đường



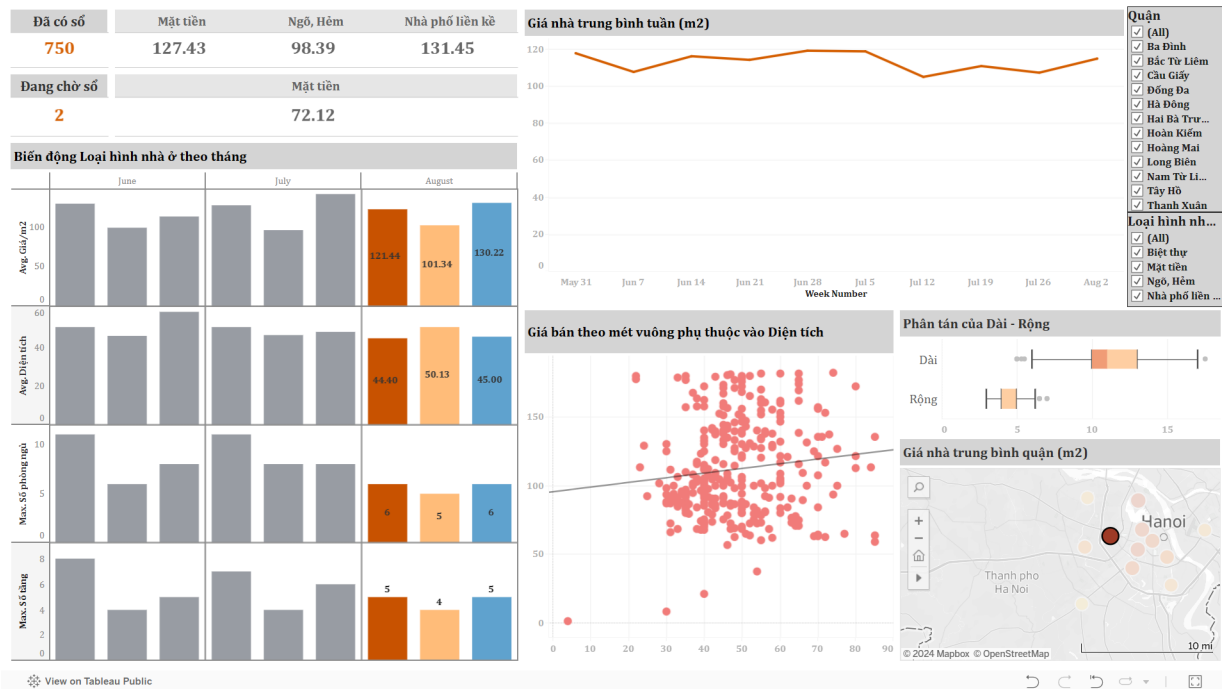
Hình 4.13 Tương tác với biểu đồ đường

Như đã được đề cập trước đó, trong biểu đồ đường, có sẵn một bảng điều khiển giúp lựa chọn các **Quận** và **Loại hình nhà ở** cụ thể để hiển thị trên biểu đồ, đồng thời bảng này cũng có thể được áp dụng cho các loại biểu đồ khác. Ví dụ, sau khi lựa chọn **Ba Đình** và **Bắc Từ Liêm** để hiển thị, ta có thể thấy sự thay đổi tương ứng trong các biểu đồ khác. Đặc biệt, biểu đồ địa lý sẽ chỉ hiển thị thông tin của hai quận đã chọn, giúp dễ dàng quan sát và so sánh các thông số cụ thể.

**Nhận xét:** Sử dụng hình thức tương tác có thể nâng cao khả năng linh hoạt và hiểu biết sâu sắc về dữ liệu của người dùng, đặc biệt là thông tin chi tiết về từng quận. Điều này giúp người dùng xác định được vị trí của các căn nhà có giá trị và tiềm năng tốt nhất trong từng khu vực.

**Lý do áp dụng:** Theo như ta được biết, giá nhà bị ảnh hưởng bởi rất nhiều yếu tố đặc biệt như vị trí địa lý, loại hình nhà ở. Với việc áp dụng bộ lọc này vào biểu đồ ta có thể chỉ định cụ thể những giá trị thuộc tính cụ thể mà ta muốn, từ đó dashboard sẽ hiển thị thông tin theo các sự lựa chọn này. Đặc biệt hơn ở biểu đồ phân tán, khi để ở giá trị mặc định ta thấy biểu đồ này không thể hiện quá nhiều ý nghĩa khi có quá nhiều điểm dữ liệu cũng như đường hồi quy nằm ngang. Điều này có thể giải thích thông qua với một diện tích, giá nhà ở loại hình như biệt thự sẽ khác với giá nhà ở loại hình nhà trong ngõ, hẻm. Do đó việc áp dụng bộ lọc này là một bước khá cần thiết cho việc trực quan để nhìn và mang lại nhiều thông tin hơn.

### 4.3 Tương tác với biểu đồ địa lý



**Hình 4.14** Tương tác với biểu đồ địa lý

Mục đích của biểu đồ địa lý là cho ta biết được sự khác biệt giữa các vị trí dữ liệu với nhau. Do đó việc hiểu rõ hơn một điểm dữ liệu bất kỳ là điều cực kỳ cần thiết và nên được tích hợp vào dashboard. Từ đó sự tương tác với biểu đồ địa lý đã được thêm vào. Khi người dùng muốn xem xét kĩ hơn về các giá trị của một vị trí cụ thể, người dùng chỉ đơn giản chọn điểm vị trí muốn xem và sau đó dashboard sẽ thay đổi theo điểm dữ liệu này. Để dễ dàng chọn đúng vị trí, biểu đồ địa lý có thể phóng to, thu nhỏ và di chuyển theo ý thích của người dùng.

**Nhận xét:** Sử dụng sự tương tác này có thể nâng cao sự linh hoạt trong việc phân tích dữ liệu. Hơn thế nữa người dùng có thể kết hợp nhiều sự tương tác giữa các biểu đồ với nhau như kết hợp chọn 1 điểm dữ liệu ở biểu đồ địa lý và 1 cột ở biểu đồ cột. Chính điều này giúp cho người dùng có thể tùy biến và tương tác với dashboard một cách tự do, linh hoạt và thỏa mái đem đến những trải nghiệm tốt nhất cho người sử dụng.

## 5 Kết luận

Bản báo cáo này thể hiện sức mạnh của trực quan hóa dữ liệu trong việc phân tích và hiểu biết sâu sắc về thị trường bất động sản tại Hà Nội. Qua các bước từ tiền xử lý dữ liệu đến phân tích thống kê định tính và định lượng, nhóm đã xây dựng nên một bộ dashboard đa dạng, từ biểu đồ cột, biểu đồ đường, đến biểu đồ phân tán và địa lý, mỗi loại phục vụ một mục đích riêng biệt nhằm đem lại cái nhìn toàn diện về đặc trưng và xu hướng của dữ liệu. Các tương tác được tích hợp vào

dashboard cung cấp khả năng điều chỉnh và phân tích linh hoạt, cho phép người dùng khám phá dữ liệu một cách sâu sắc hơn.

## 6 Đánh giá

Nhóm đã hoàn thành tốt trong việc áp dụng các kỹ thuật trực quan hóa dữ liệu để giải quyết vấn đề cụ thể. Việc sử dụng Tableau để tạo dashboard cho phép trình bày thông tin một cách trực quan và dễ hiểu. Tuy nhiên, dự án cũng gặp một số hạn chế như khả năng của dữ liệu chưa thể hiện đầy đủ mọi khía cạnh của thị trường bất động sản, chẳng hạn như các yếu tố vĩ mô có thể ảnh hưởng đến giá nhà. Ngoài ra, mặc dù các tương tác trên dashboard rất hữu ích nhưng cần đảm bảo rằng chúng không làm phức tạp hóa trải nghiệm người dùng, đặc biệt là đối với những người không quen với việc phân tích dữ liệu.

Đánh giá dựa trên các tiêu chí:

- Kết hợp nguồn dữ liệu đáng tin cậy.
- Phù hợp với mục đích.
- Rõ ràng và dễ hiểu.
- Phân tích được sự thay đổi và xu hướng.
- Tương tác và điều hướng.
- Thiết kế hấp dẫn.
- Tính phản hồi.
- Hiệu suất.