

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH



Group 5: Final Data Visualization Report

1. Thông tin nhóm

MSSV	Tên	Công việc	Đánh giá
19127587	Trương Chí Toàn	Tác động các biến khác đến giá nhà	Hoàn thành tốt công việc
19127149	Trần Việt Hoàng	Thống kê biểu đồ khu vực	Hoàn thành tốt công việc
19127592	Lê Minh Trí	Chuẩn bị tập dữ liệu + Phân tích + Dự đoán	Hoàn thành tốt công việc
18127248	Võ Trần Quang Tuấn	Chọn lọc các tiêu chí chung khi mua nhà	Hoàn thành tốt công việc
18127266	Trần Quỳnh Như	Biến động giá nhà qua thời gian	Hoàn thành tốt công việc

2. Giới thiệu về tập dữ liệu

Đây là một tập dữ liệu tổng quan về giá nhà với mỗi dòng là một nhà khác nhau với thời điểm đăng ký khác nhau. Tập dữ liệu có các biến chính sau:

- Ngày: Số ngày đăng ký.
- Địa chỉ: Địa chỉ chính xác nhà được đăng ký bán, đã bao gồm quận và huyện.
- Quận: Tên quận.
- Huyện: Tên Huyện của từng quận.
- Loại hình nhà ở: Bao gồm 4 loại hình khác nhau, “Nhà ngõ hẻm”, “Nhà mặt phố, mặt tiền”, “Nhà phố liền kề”, “Nhà biệt thự”.
- Giấy tờ pháp lý: “Đã có sổ”, “Đang chờ sổ”, “Giấy tờ khác”.
- Số tầng: số lượng tầng,
- Số phòng: số lượng phòng.
- Diện tích: Diện tích được tính theo chiều dài rộng.
- Dài, Rộng: Chiều dài và chiều rộng.
- Giá/m2: Giá nhà với mỗi mét vuông.

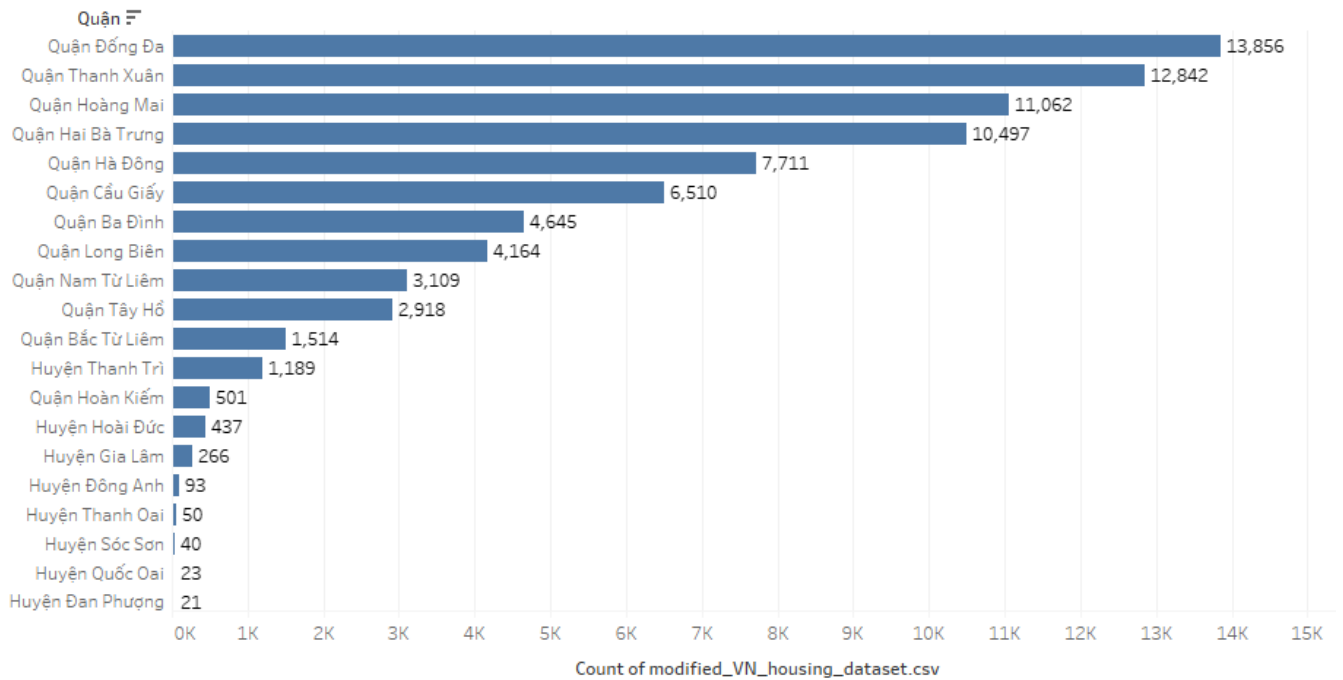
Bằng cách trực quan hóa cũng như phân tích, ta có thể rút ra được các biến cũng như ý nghĩa quan trọng. Từ đó trích xuất được các thông tin hữu ích cho việc mua nhà của khách hàng, giúp họ thuận lợi hơn trong việc chọn các căn nhà phù hợp với túi tiền cũng như nhu cầu.

3. Trực quan hóa

3.1. Các tiêu chí chung cần chọn khi mua nhà

Mỗi người dân khi có nhu cầu mua nhà cần phải xem xét các tiêu chí trong dữ liệu nhà ở được bán. Để chọn được một ngôi nhà ưng ý, điều đó cần phụ thuộc vào nhiều đặc điểm, thuộc tính được thống kê trong dữ liệu có sẵn. Do đó, chúng ta cần nghiên cứu và trực quan hóa dữ liệu dựa theo các thuộc tính trong tập dữ liệu để đưa ra các phán đoán và nhận xét chung cho người cần mua nhà. Chúng ta sẽ làm rõ thông qua các phần trực quan và nhận xét các biểu đồ cột.

3.1.1. Trực quan phân bố nhà theo từng quận



Sử dụng biểu đồ cột để trực quan dữ liệu tương ứng với cột “Quận”, chúng ta rút ra được một số nhận xét sau:

- Dữ liệu phân bố không đều giữa các quận, có sự chênh lệch khá lớn và rõ rệt.
- Nhìn vào biểu đồ, ta thấy được Quận Đống Đa là nơi có nhiều nhà nhất (13856 nhà).
- Quận Thanh Xuân có số nhà không kém gì quận Đống Đa (12842 nhà), cũng được tính trong các quận có số nhà khá cao.
- Chúng ta có thể liệt kê được một số quận có số nhà cao, như Quận Đống Đa (13856), Quận Hai Bà Trưng (10497), Quận Hoàng Mai (11062), Quận Thanh Xuân (12842).
- Huyện Ba Vì là nơi có ít nhà bán nhất (5 nhà).
- Nhìn chung, dữ liệu phân bố không đều, từ biểu đồ cho thấy nhà có xu hướng tập trung nhiều ở các quận lớn, thưa thớt ở các huyện vùng xa.

Khi có nhu cầu mua nhà, chúng ta sẽ chọn nhà dựa trên các tiêu chí:

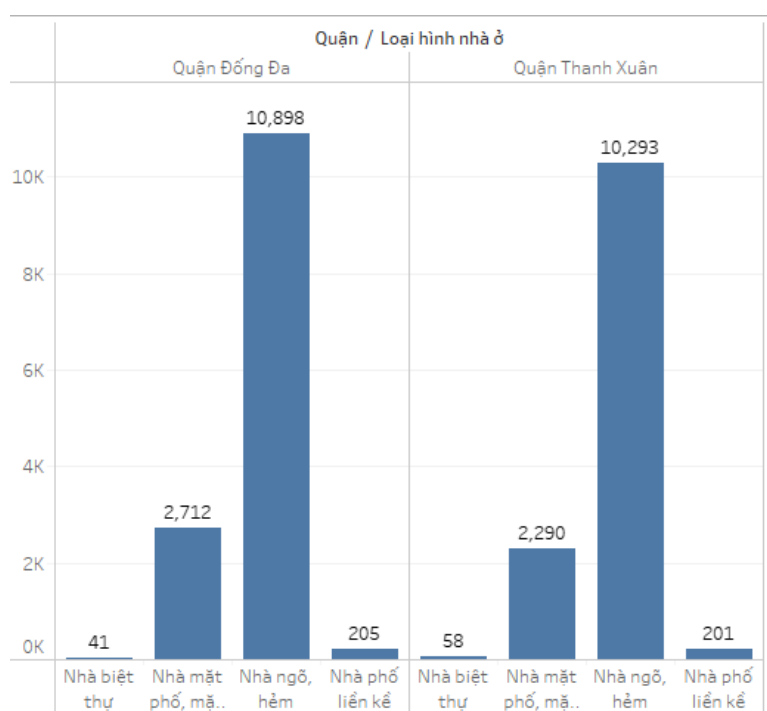
- Nhà ở khu có số lượng nhà bán cao, vì điều kiện sống sẽ tốt hơn, thích hợp cho những người yêu thích sự sầm uất của đô thị.
- Nhà ở khu có số lượng nhà bán thấp, vì không khí sẽ yên tĩnh hơn, thích hợp cho công nhân viên chức nhà nước, những người yêu thích sự yên tĩnh, không sầm uất.

Để xem xét nên chọn nhà như thế nào, chúng ta sẽ xem xét tiếp sự phân bố các loại hình nhà ở trong những quận được chọn theo hai tiêu chí trên. Lọc ra các quận theo hai tiêu chí:

- Khu có số lượng nhà bán cao: Quận Đống Đa, Quận Thanh Xuân

- Khu có số lượng nhà bán thấp: Quận Nam Từ Liêm, Quận Tây Hồ

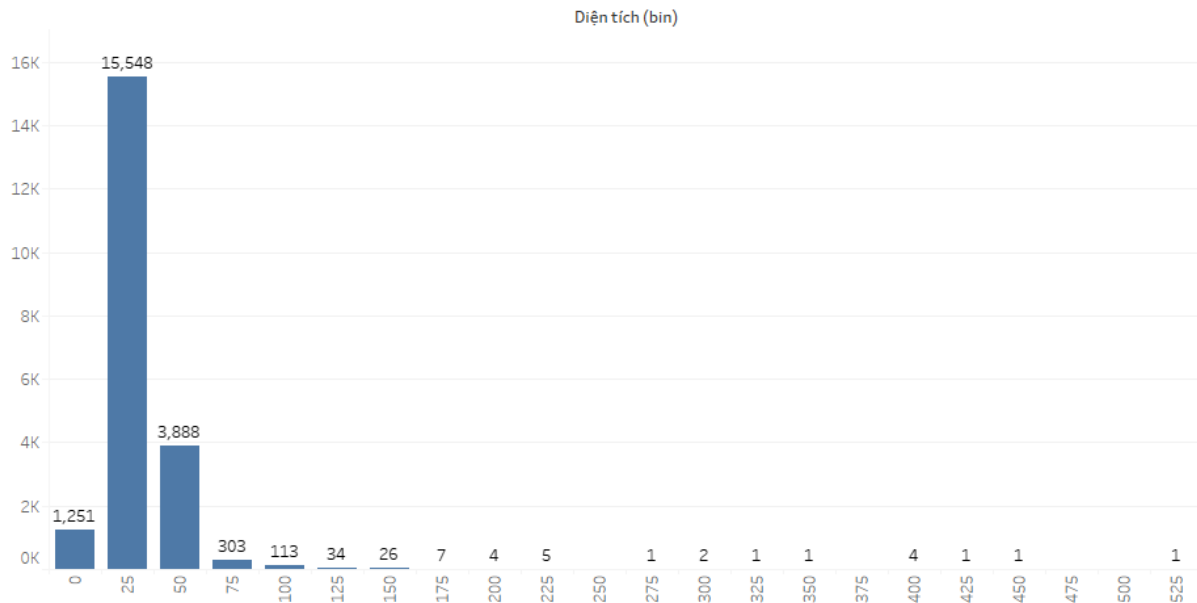
3.1.2. Trực quan hóa dữ liệu nhà tại khu có số lượng nhà bán cao



Chúng ta xét Quận Đống Đa và Quận Thanh Xuân. Trực quan thuộc tính “Loại hình nhà ở” cho hai quận này, ta được biểu đồ cột như trên, và có thể rút ra được một số các nhận xét:

- Nhà ngõ, hẻm chiếm số lượng lớn trong loại hình nhà ở
- Nhà biệt thự chiếm số lượng ít nhất

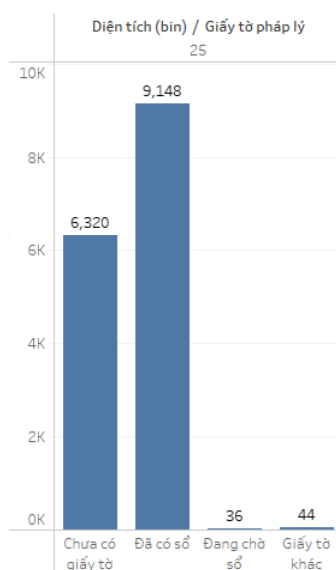
Điều đó cho thấy, ở khu có số lượng nhà bán cao, thì loại hình nhà biệt thự không phải là xu hướng, thay vào đó, loại hình nhà ngõ, hẻm là ưu tiên cho người dân. Để xem xét về diện tích của loại nhà ngõ, hẻm tại đây, chúng ta cần trực quan thêm thuộc tính diện tích.



Bằng phương pháp chia giỏ để xử lý cho dữ liệu diện tích, chúng ta trực quan để thấy được phân phối các diện tích nhà loại nhà ngõ, hẻm và rút ra được một số nhận xét:

- Dữ liệu ngoại lai khá nhiều, các dữ liệu không ngoại lai tập trung chính vào 3 giá trị:
 - Nhà có diện tích từ 0 -> 25 m² (Diện tích loại 1)
 - Nhà có diện tích từ 25m² -> 50m² (Diện tích loại 2)
 - Nhà có diện tích từ 50m² -> 75m² (Diện tích loại 3)
- Chúng ta có thể thấy được nhà có diện tích loại 2 được ưu tiên chọn lựa hơn (với 15548 bản ghi thống kê).
- Do đó, người dân có mong muốn chọn các ngôi nhà diện tích loại 2 hơn.

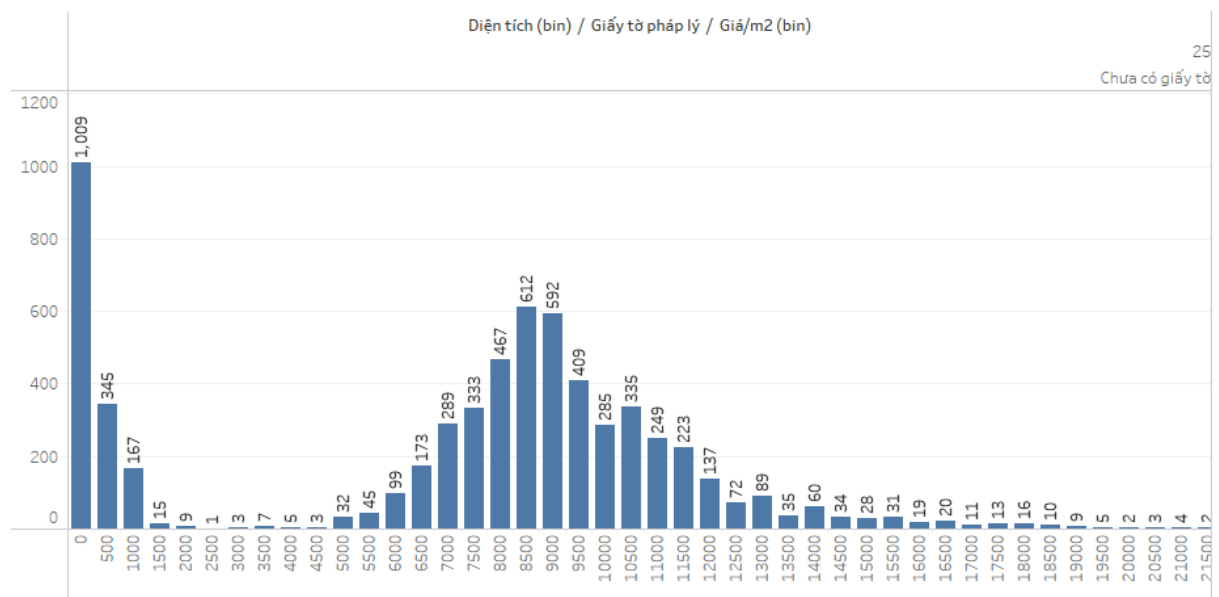
Xem xét thêm về mặt giấy tờ pháp lý của các ngôi nhà diện tích loại 2, chúng ta trực quan được biểu đồ sau:



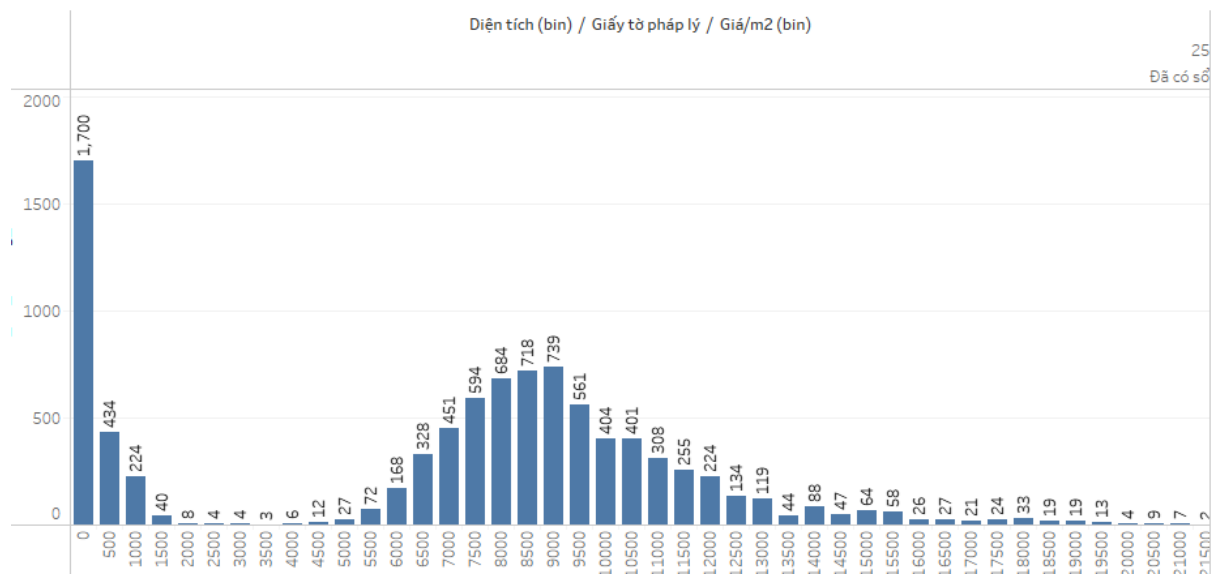
Quan sát kĩ biểu đồ, chúng ta sẽ ưu tiên hơn cho việc chọn hai loại giấy tờ pháp lý: chưa có giấy tờ hoặc đã có sổ. Trực quan thêm về mặt giá nhà để đưa ra quyết định hợp lý nhất khi mua nhà, ta

được:

Đối với “Chưa có giấy tờ”:



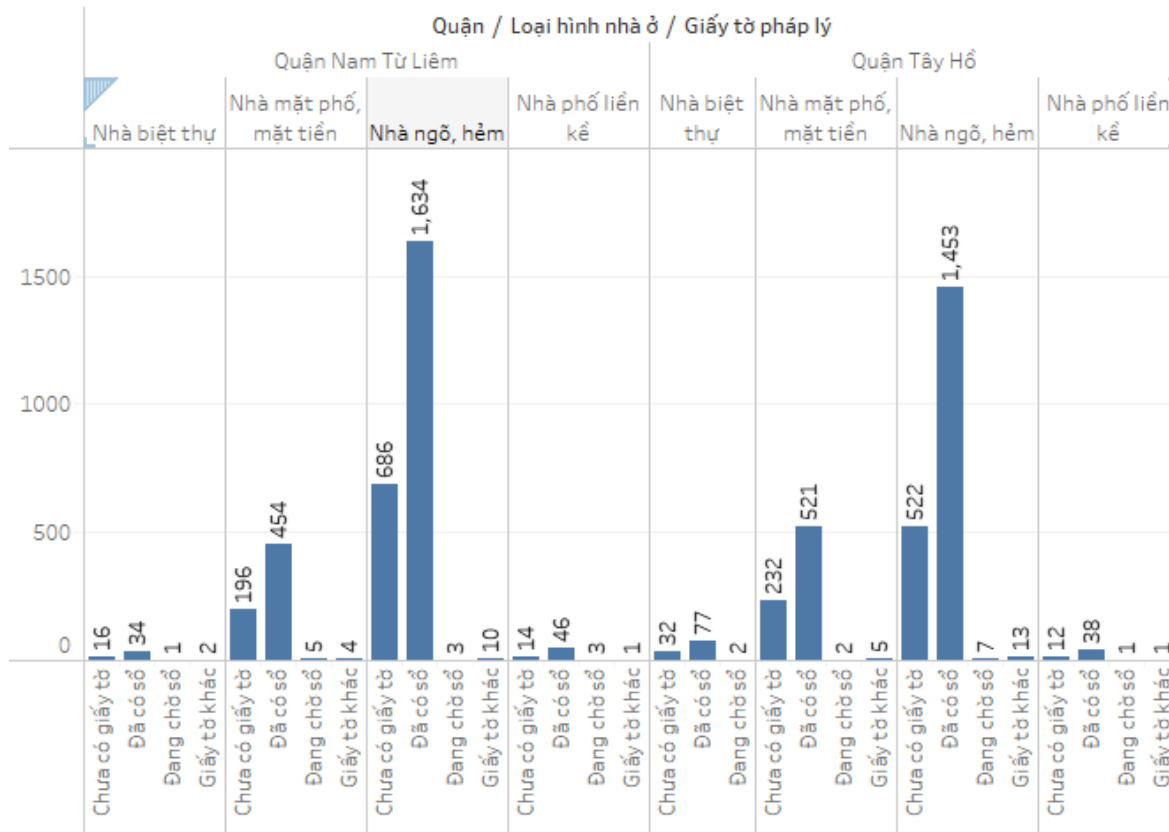
Đối với “Đã có sổ”:



Nhìn chung, hai biểu đồ có hình dáng tương đương nhau, cho thấy:

- Dù trong hai loại giấy tờ pháp lý khác nhau, nhưng giá nhà vẫn có sự tương đồng khi phân tích (hình dạng của biểu đồ)
- Cho thấy sự phân biệt rõ rệt giữa hai loại giá nhà:
 - Những người không có nhiều điều kiện sẽ chọn giá nhà thấp (0 -> 500/m2), cộng đồng này chiếm số đông trong những người mua nhà tại khu có số lượng nhà bán cao.
 - Những người có nhiều điều kiện hơn một chút sẽ mua nhà có giá tăng dần trong khoảng 7500 -> 10000/m2

3.1.3. Trục quan hóa dữ liệu nhà tại khu có số lượng nhà bán thấp



Xem xét kĩ biểu đồ được trục quan cho quận Nam Từ Liêm và quận Tây Hồ, kèm theo các tiêu chí loại nhà và giấy tờ pháp lý, chúng ta thấy được:

- Có sự tương đương giữa phân bố nhà ở khu có số lượng nhà bán cao và khu có số lượng nhà bán thấp
- Loại hình nhà ở “nhà ngõ, hẻm” vẫn chiếm số đông trong khu có số lượng nhà bán thấp
- Người dân tại khu dân cư thưa vẫn có nhu cầu chọn nhà “chưa có giấy tờ” hoặc “đã có sổ”.

3.1.4. Nhận xét chung về các tiêu chí mua nhà

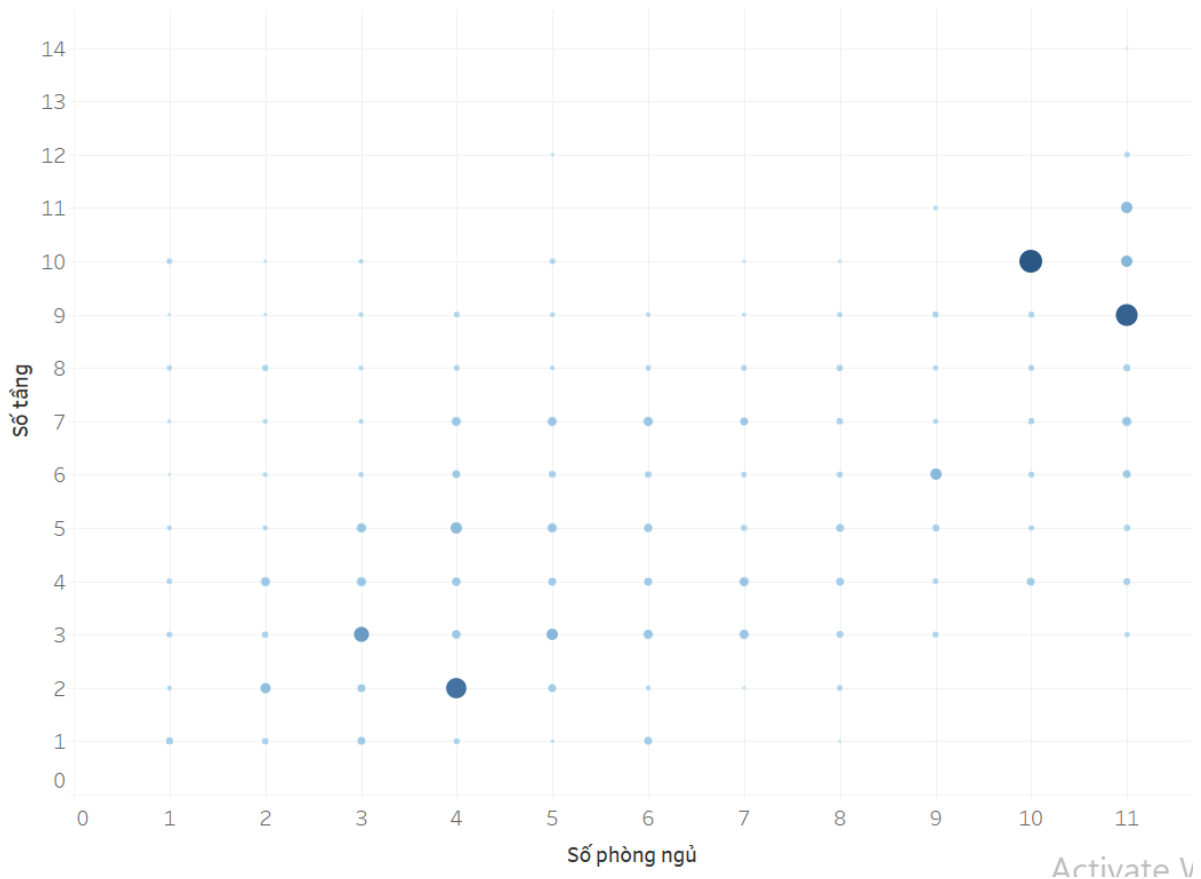
Từ việc tách hai loại khu dân cư dựa vào phân bố nhà theo quận và trục quan hóa các thuộc tính, ta có thể rút ra được:

- Bởi vì ta đang xét trên hai tiêu chí đối lập nhau (quận có số lượng nhà bán cao và quận có số lượng nhà bán thấp), thông qua các biểu đồ trên ta thấy việc phân bố nhà dựa theo quận không gây nên sự khác biệt về nhu cầu mua nhà và tiêu chí chọn nhà của người dân.
- Loại nhà có giấy tờ pháp lý “đã có sổ”, hoặc “chưa có giấy tờ” vẫn được ưu tiên hơn hai loại giấy tờ còn lại.
- Nhà ngõ, hẻm là lựa chọn của hầu hết các hộ gia đình.
- Về giá cả, có sự phân hóa rõ ràng giữa hai loại “gia cảnh”:
 - Gia cảnh bình thường: thường chọn nhà có giá từ 0 -> 500/m²
 - Gia cảnh có điều kiện: thường chọn nhà có giá từ 7500 -> 10000/m²

3.2. Mối tương quan giữa các cột số tầng, số phòng ngủ cũng như diện tích ảnh hưởng đến giá/m².

Ở đây ta vẽ biểu đồ scatter plot với trục x là số phòng ngủ, trục y là số tầng, chấm có màu càng đậm, kích thước càng lớn nghĩa là nhà ấy có diện tích càng lớn.

Vì có những căn nhà có số tầng rất lớn (lớn nhất là 73 tầng), thường những căn nhà đó sẽ là khách sạn hay chung cư, đồng thời khiến cho những việc trực quan rất là xấu, nên em sẽ giới hạn số tầng là 20 để tránh các outlier



Như ta thấy thì số lượng phòng ngủ và số tầng được phân bố trải đều theo đường chéo. Tức là tỉ lệ số phòng ngủ sẽ tỉ lệ thuận với nhau, càng nhiều tầng thì sẽ có càng nhiều phòng ngủ.

Như ở trên đã phân tích thì những chấm nhỏ xuất hiện nhiều và phân bố đều, đó là những nhà ngõ, hẻm, chiếm phần lớn như đã phân tích ở trên.

Có 4 điểm nổi bật hơn các điểm còn lại ở trên biểu đồ. Vì có diện tích lớn nhất nên khả năng cao là sẽ có nhà biệt thự ở trong ấy. Ta sẽ phân tích rõ ràng hơn về loại hình nhà ở, quận huyện, diện tích và giá/m² của 4 căn nhà này.

Loại hình nhà ở	Quận, Huyện	Diện tích	Giá/m ²
Nhà mặt phố, mặt tiền	Huyện Gia Lâm	1154	607
Nhà biệt thự	Huyện Quốc Oai	1900	842
Nhà mặt phố, mặt tiền	Quận Thanh Xuân	2400	1875
Nhà mặt phố, mặt tiền	Quận Cầu Giấy	2200	723

Như vậy 4 căn trên thì có 3 căn là nhà mặt phố, 1 căn là nhà biệt thự. 2 căn ở huyện và 2 căn ở quận. Diện tích 2 căn ở huyện trung bình thấp hơn so với 2 căn ở quận. Còn về giá cả thì nhà ở quận trung bình lớn hơn so với nhà ở huyện.

Qua bảng trên, ta rút ra được các nhận xét:

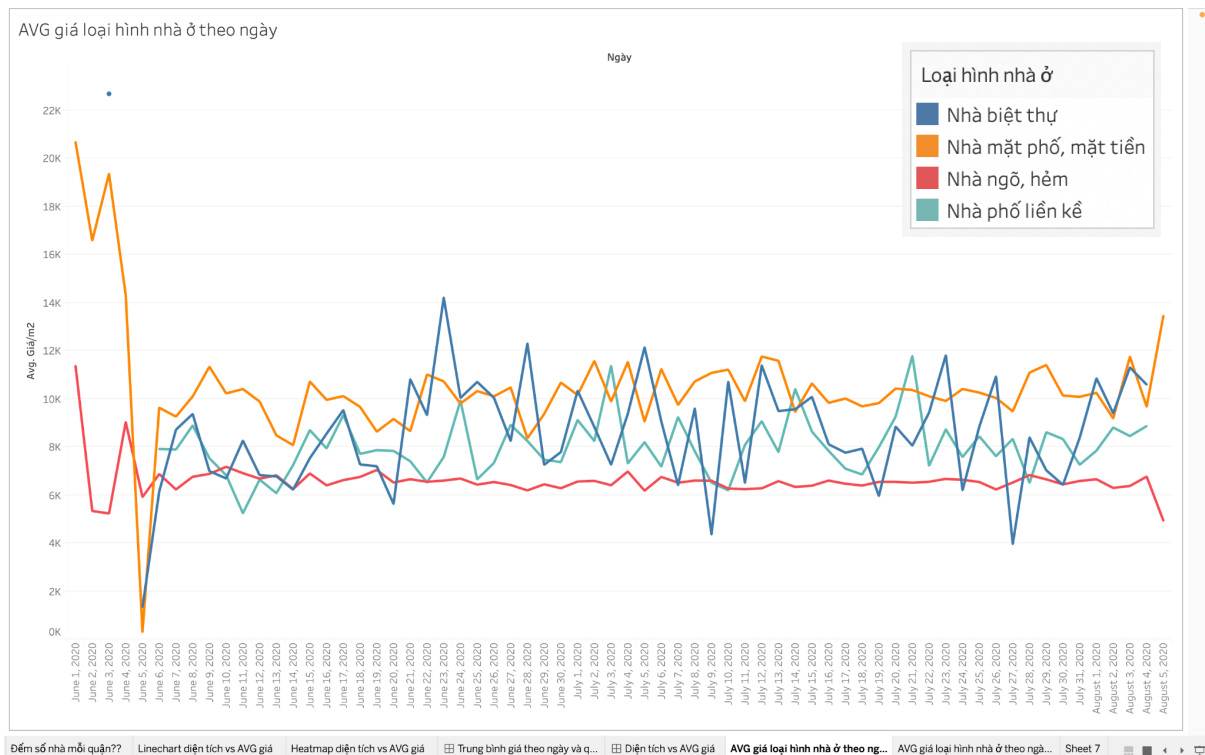
- Những căn nhà có diện tích lớn hầu hết sẽ là những loại nhà mặt phố, mặt tiền hoặc biệt thự
- Nhà có diện tích càng lớn thì đi kèm theo giá sẽ càng cao
- Tuy nhiên, giá/m² thì sẽ phụ thuộc vào vị trí địa lý, như là ở quận thì sẽ có giá thành cao hơn ở huyện. Nhà biệt thự ở huyện giá còn thấp hơn 1 vài căn nhà ở quận.

Tùy vào nhu cầu của khách hàng, mà ta sẽ giới thiệu cho họ những căn nhà khác nhau.

- Khách muốn những căn nhà rộng rãi, nhiều tầng mà giá thành rẻ thì ta sẽ giới thiệu những căn nhà ở ngoại thành
- Còn nếu khách muốn những căn nhà ở gần trung tâm để thuận tiện cho việc đi học, đi làm với giá cả phải chăng thì ta sẽ giới thiệu những căn nhà có diện tích nhỏ, ít tầng để phù hợp với túi tiền của khách hàng.

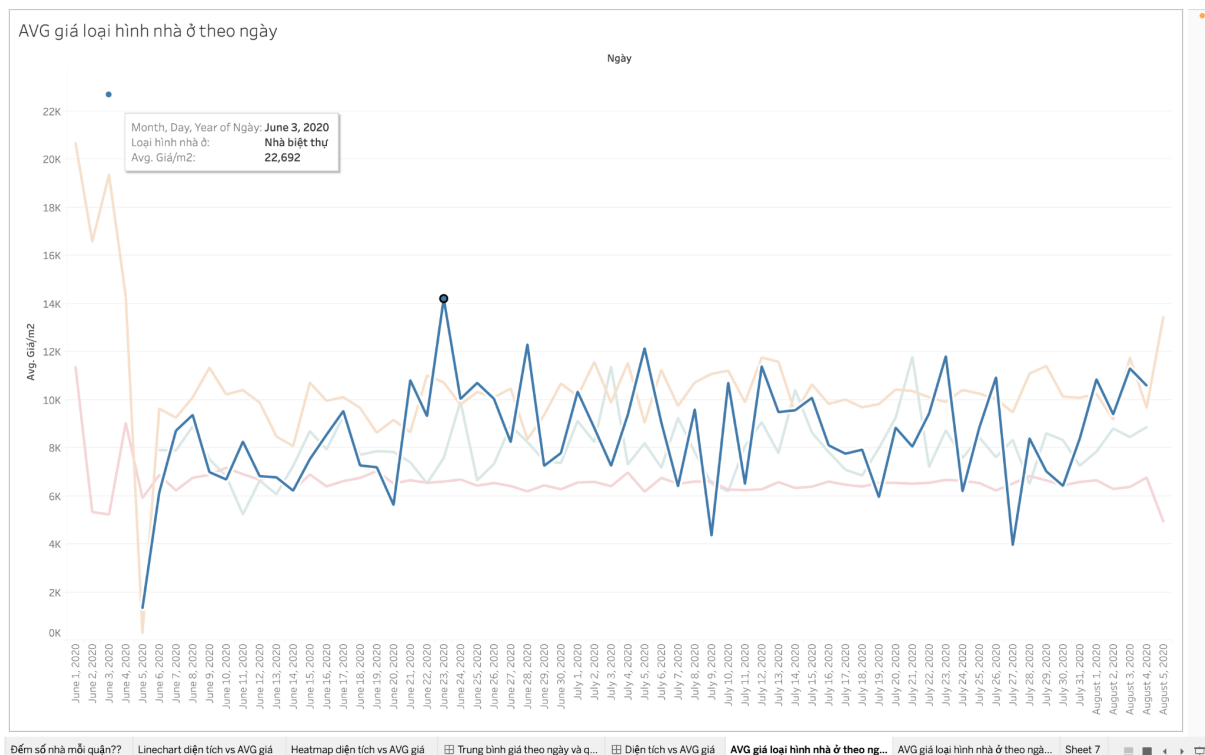
3.3. Biến động giá nhà trên các tiêu chí khác nhau

Sau đây là biểu đồ thể hiện **trung bình giá nhà** của **4 loại hình nhà ở** khác nhau qua từng ngày từ tháng 6/2020 đến tháng 8/2020.

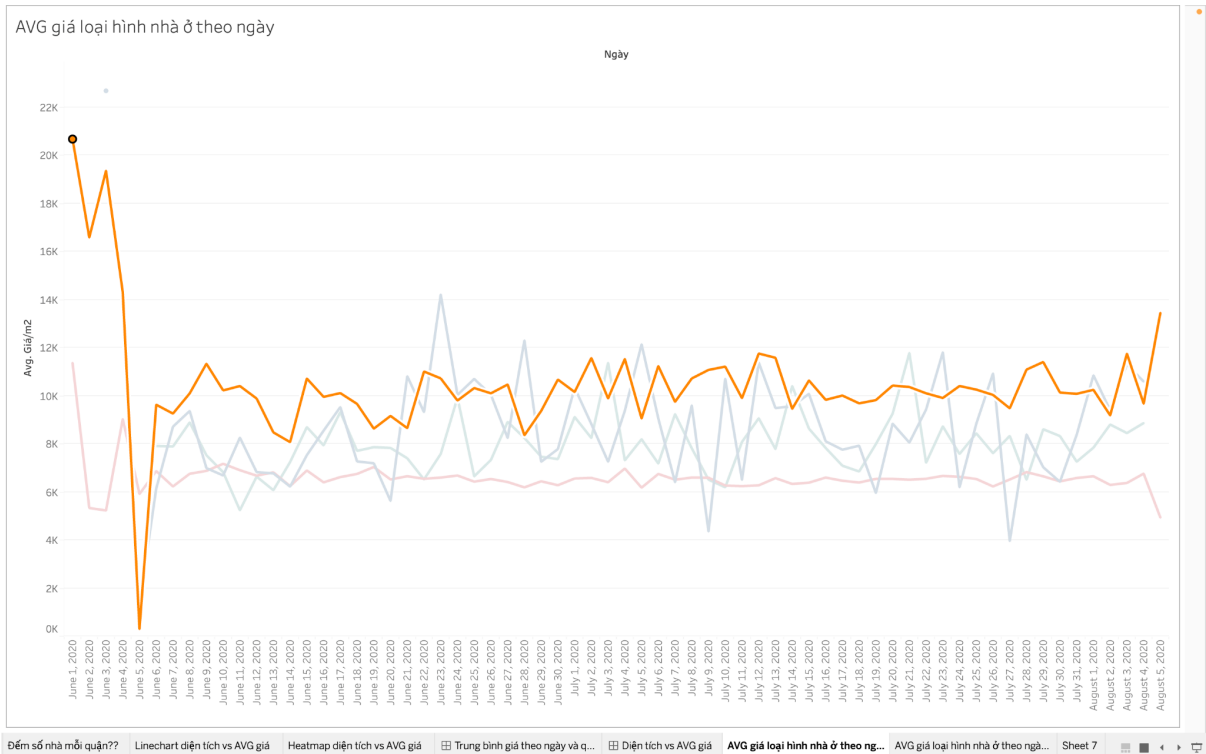


Sau đây ta sẽ thực hiện phân tích dữ liệu của từng loại hình nhà ở:

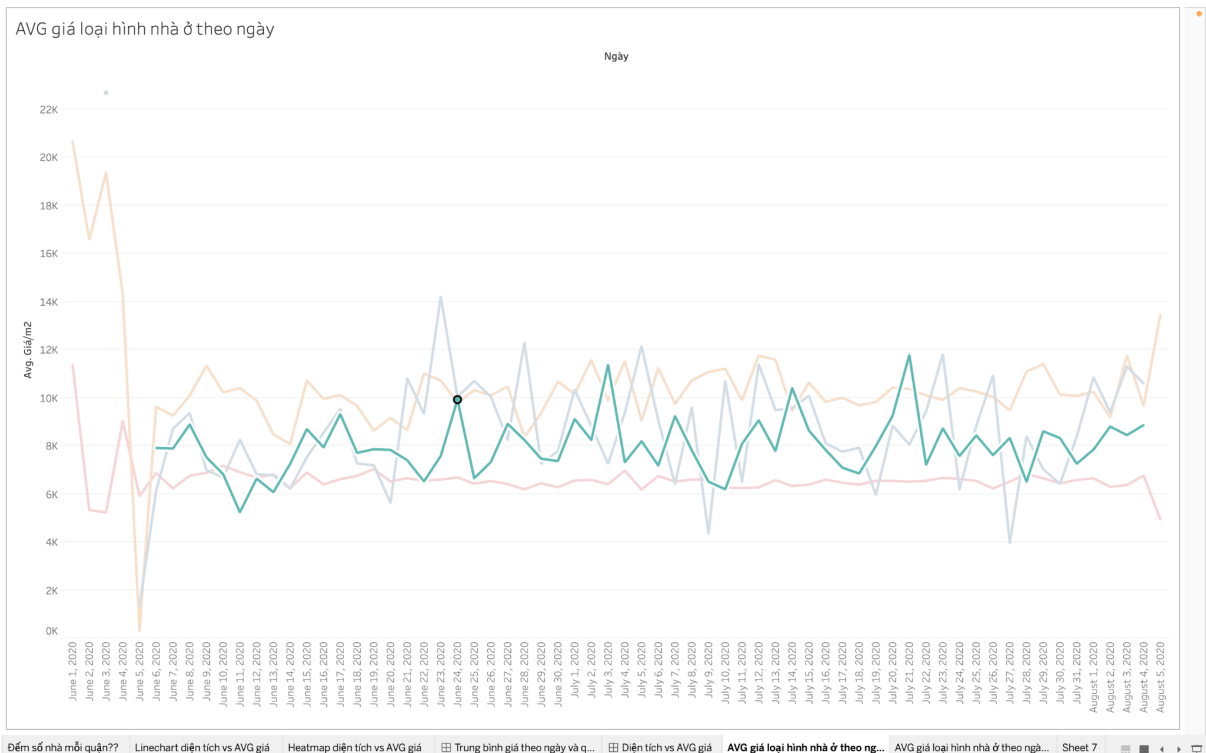
- **Nhà biệt thự:** Nhìn chung, đây là loại hình nhà ở có *giá trị biến động mạnh nhất*. Vào ngày 03/06/2020, giá nhà biệt thự đạt giá trị cao nhất trong tất cả 4 loại hình và trong suốt khoảng thời gian 3 tháng. Tuy nhiên sau đó, giá biệt thự giảm mạnh vào ngày 05/06/2020, sau đó có chiều hướng tăng lên cao rồi lại giảm xuống thấp bất ổn định.



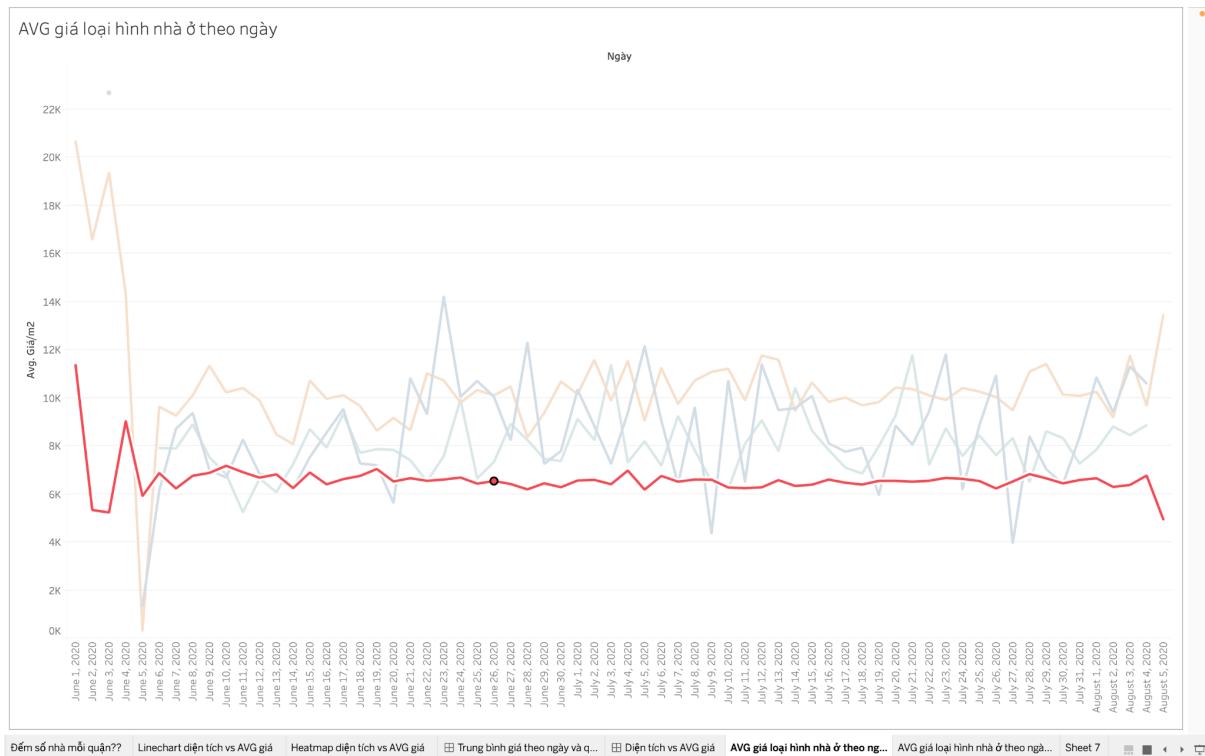
- Nhà mặt phố, mặt tiền:** đây là loại hình nhà ở có *giá trị trung bình khá cao*, tuy nhiên *giá trị lại khá ổn định theo thời gian*. Mặc dù vào đầu tháng 6, loại hình nhà này có giá trị cao ngất và sau đó tụt dốc không phanh, đến ngày 06/06/2020 thì giá nhà mặt phố, mặt tiền đã tăng lên trở lại và giữ ở mức ổn định, chỉ tăng hoặc giảm nhẹ.



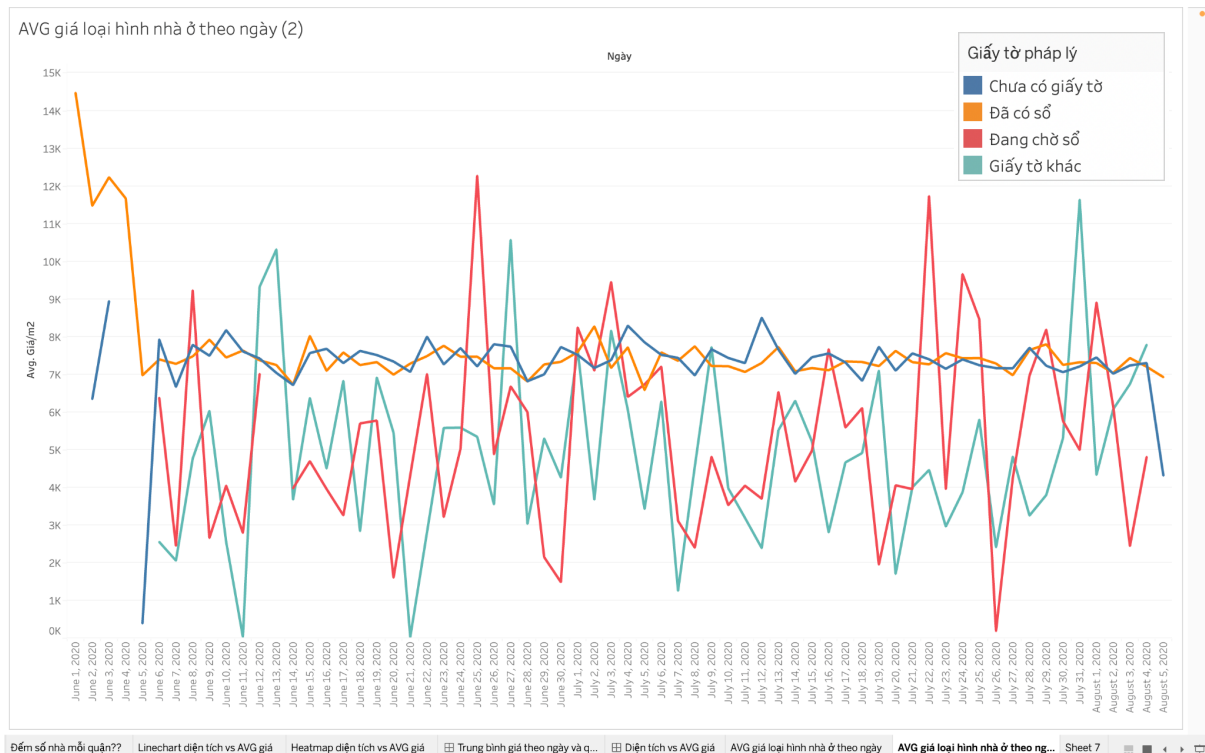
- Nhà phố liền kề:** có giá trị *biến động trung bình*, và nhìn chung giá trị trung bình cũng thấp hơn nhà biệt thự và nhà phố liền kề.



- Nhà ngõ, hẻm: Đây là loại hình nhà ở có *giá trị trung bình thấp nhất* trong 4 loại, và đây cũng là loại hình có *giá trị ổn định nhất* qua từng ngày. Vào đầu tháng 6, tuy giá nhà ngõ hẻm khá cao, nhưng sau đó giá đã giảm và hầu như là giữ nguyên, chỉ tăng hoặc giảm rất nhẹ.



Tiếp theo, ta thực hiện phân tích biểu đồ thể hiện **trung bình giá nhà** của **4 loại giấy tờ pháp lý** khác nhau qua từng ngày từ tháng 6/2020 đến tháng 8/2020.



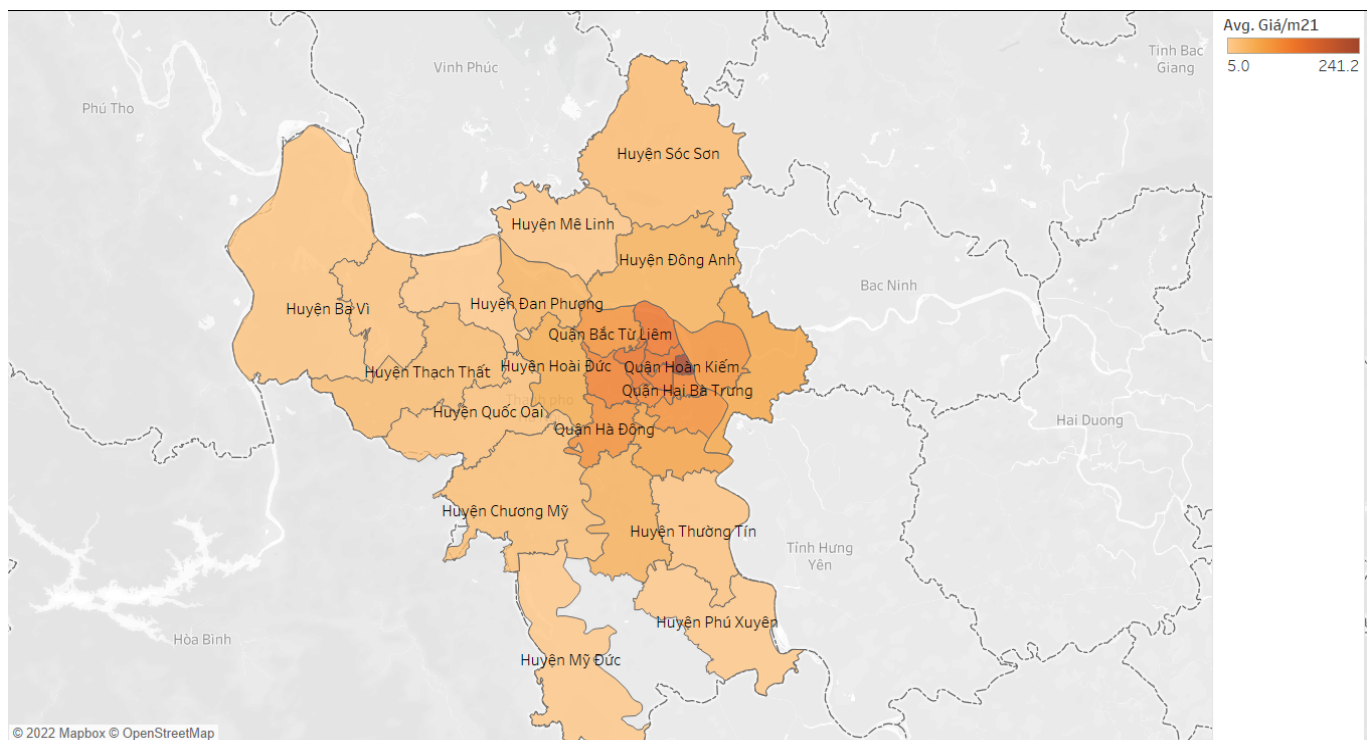
Nhìn chung, ta có thể nhận thấy biểu đồ này được chia thành 2 nhóm khác nhau:

- Đã có sổ và chưa có giấy tờ: đây là 2 loại nhà có *giá trị trung bình khá cao*, tuy nhiên lại *rất ổn định* qua từng ngày.
- Đang chờ sổ và giấy tờ khác: đây là 2 loại nhà có *giá trị biến đổi bất ổn nhất*, có thời điểm giá trị trung bình rất cao, nhưng sau đó lại giảm cũng rất mạnh.

Sau khi phân tích giá trung bình qua từng ngày của 4 loại hình nhà ở và 4 loại giấy tờ pháp lý, ta có thể rút ra các kết luận như sau:

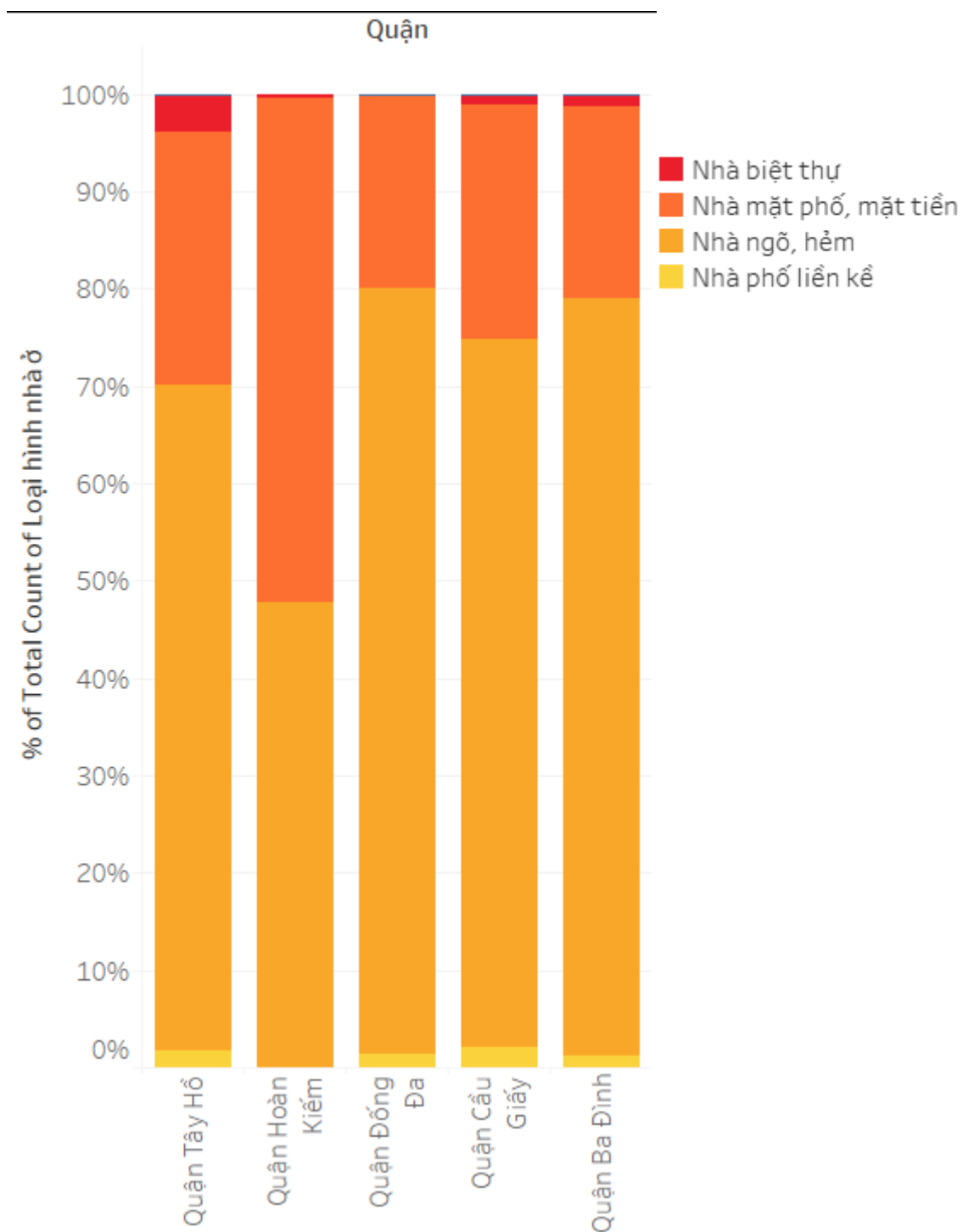
- Đối với những đối tượng muốn mua nhà với mức giá ổn định, ta có thể khuyên họ nên mua:
 - Nhà mặt phố, mặt tiền hoặc nhà ngõ hẻm, đã có sổ hoặc chưa có sổ: vì đối với các thuộc tính này, giá trị trung bình của nhà thường rất ổn định, nếu có biến động thì giá cũng tăng hoặc giảm nhẹ.
- Đối với những đối tượng mua nhà phục vụ cho mục đích khác như kinh doanh, bất động sản, ta có thể tư vấn họ mua:
 - Nhà biệt thự hoặc nhà phố liền kề, đang chờ sổ và giấy tờ khác: vì đây là những loại nhà có giá trị biến động mạnh, thuận tiện cho việc mua đi bán lại. Người mua có thể lựa chọn thời điểm giá nhà thấp, sau đó qua các biểu đồ trên, họ có thể suy ra thời điểm giá nhà tăng để có thể thu lại lợi nhuận cao nhất.

3.4. Thống kê biểu đồ khu vực



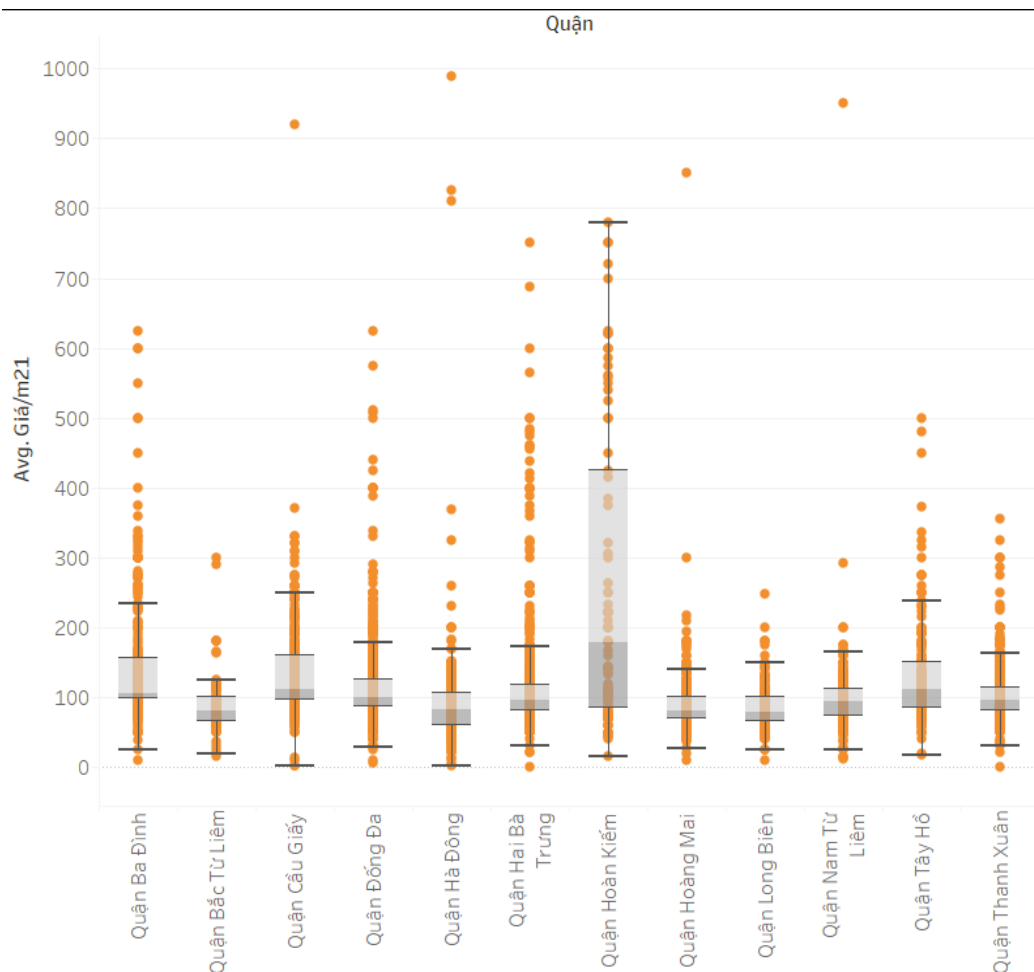
- Nơi có trung bình tỷ lệ giá/m² cao nhất là quận Hoàn Kiếm với trung bình 241.2 triệu/m². Thấp nhất là huyện Ba Vì với 11.0 triệu/m². Xét trong khu vực nội thành thì thấp nhất là quận Bắc Từ Liêm với 84.3 triệu/m²
- Quận Hoàn Kiếm có trung bình giá nhà trên m² gần như gấp đôi quận đứng thứ 2.
- Nhìn chung, giá nhà trên m² ở khu vực ngoại thành tương đối đồng đều. Giá nhà có xu hướng tăng càng về nội thành.

Ta sẽ tập trung thống kê 5 khu vực nội thành có màu đỏ đậm nhất.



Biểu đồ thể hiện cơ cấu loại nhà ở của top 5 quận về giá/m².

- Loại hình nhà ở phổ biến là loại hình nhà ngõ, hẻm. Thứ 2 là loại hình nhà mặt phố.
- Quận Hoàng Mai có phần trăm số lượng nhà ngõ, hẻm lớn nhất (82.5%). Quận Hoàn Kiếm có phần trăm nhà biệt thự (3.9%) và nhà mặt phố (51.89%) lớn nhất.
- Quận hoàn kiếm có tỉ lệ gần như 50/50 giữa nhà mặt phố và nhà ngõ hẻm. Đáng chú ý là quận Hoàn Kiếm có giá nhà trên m² cao vượt trội so với các khu vực còn lại. Quận Hoàn Kiếm cũng không có phân bố nhà biệt thự đáng kể (0.38%) và có diện tích nhỏ nhất trong tất cả các quận.
- Nhìn chung, sự phân bố giữa các loại hình nhà không có quy luật thống nhất.



- Dựa theo biểu đồ trên, ta có thể thấy Quận Hoàn Kiếm đáng chú ý nhất có khoảng giá trị cao nhất cũng như dữ liệu bị skew về bên trái. Có một điểm đáng chú ý là các outlier là không có, cùng lắm các điểm dữ liệu tập trung về hai phía
- Tiếp theo, ta có các quận Ba Đình, Cầu Giấy, Tây Hồ và Thanh Xuân có xu hướng cân bằng tức không bị nghiêng về bất cứ bên nào. Tuy nhiên các quận này có điểm chung là sự chênh nhau giữa các giá trị trong khoảng các outlier là khá cao, các outliers xuất hiện chỉ chít ở khoảng bên phải của dữ liệu.
- Các quận còn lại có giá nhà đều nằm ở mức 100-200 thấp hơn mặt bằng chung so với các quận còn lại.

