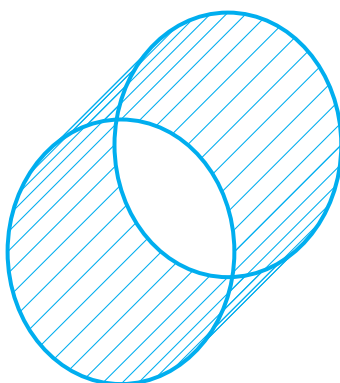


TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG - HCM
Khoa Toán - Tin Học

BÁO CÁO ĐỒ ÁN CUỐI KỲ MÔN
TRỰC QUAN HOÁ DỮ LIỆU



Khoa Toán - Tin học
Fac. of Math. & Computer Science

Giảng viên phụ trách: Nguyễn Thanh Bình
Sinh viên thực hiện: Đinh Anh Huy - 18110103
Nguyễn Đức Vũ Duy - 18110004

TP.Hồ Chí Minh - Tháng 1, 2020

Mục lục

1	Mô tả dữ liệu	3
1.1	Nội dung dữ liệu	3
1.2	Thông tin các cột của dữ liệu	3
2	Tiền xử lý dữ liệu	4
2.1	Load dữ liệu	4
2.2	Một số thông tin cơ bản của dữ liệu	4
2.2.1	In ra thông tin của dữ liệu	4
2.2.2	In ra mô tả về dữ liệu	5
2.2.3	Missing values	7
3	Phân tích dữ liệu	7
4	Trực quan hoá dữ liệu	9
4.1	Thống kê và vẽ các biểu đồ biểu diễn các features quan trọng	9
4.1.1	SalePrice	9
4.1.2	YearBuilt	9
4.1.3	YrSold	13
4.1.4	Size(sqf)	14
4.1.5	TimeToBusStop	15
4.1.6	TimeToSubway	16
4.2	Thống kê và vẽ các biểu đồ biểu diễn mối quan hệ giữa các features	18
4.2.1	Thống kê số lượng và giá bán trung bình của các căn hộ theo từng năm xây dựng	18
4.2.2	Thống kê số lượng và giá bán trung bình của các căn hộ theo từng năm bán ra	20
4.2.3	Thống kê giá giao dịch trung bình theo số lượng tiện ích có trong căn hộ.	22

4.2.4	Thống kê giá bán trên một đơn vị diện tích (square feet).	23
4.2.5	Thống kê giá giao dịch trên từng square feet theo tuổi thọ căn hộ. . .	24
4.2.6	Sự thay đổi giá giao dịch theo kích thước và năm giao dịch	25
4.2.7	Mối quan hệ giữa tuổi thọ căn hộ và giá giao dịch trên mỗi square feet theo thời gian đến trạm xe buýt gần nhất	26
4.2.8	Mối quan hệ giữa tuổi thọ căn hộ và giá giao dịch trên mỗi square feet theo thời gian đến trạm tàu điện ngầm gần nhất	27
5	Kết luận	28

1 Mô tả dữ liệu

Bộ dữ liệu đầu vào: 7_Daegu_Real_Estate_data.csv

1.1 Nội dung dữ liệu

Đây là dữ liệu giao dịch các căn hộ được tạo từ tháng 8/2007 đến tháng 8/2017 tại phường Daebong, thành phố Daegu, Hàn Quốc.

1.2 Thông tin các cột của dữ liệu

<i>SalePrice</i>	: Giá bán tính bằng đơn vị US dollar (target feature).
<i>YearBuilt</i>	: Năm xây dựng.
<i>YrSold</i>	: Năm bán.
<i>MonthSold</i>	: Tháng bán.
<i>Size(sqf)</i>	: Kích thước của căn hộ tính bằng đơn vị square feet.
<i>Floor</i>	: Số tầng của căn hộ.
<i>HallwayType</i>	: Loại phòng trước.
<i>HeatingType</i>	: Loại lò sưởi.
<i>AptManageType</i>	: Loại hình thức quản lý của căn hộ.
<i>N_Parkinglot(Ground)</i>	: Số chỗ đỗ xe ở tầng trệt.
<i>N_Parkinglot(Basement)</i>	: Số chỗ đỗ xe dưới tầng hầm.
<i>TimeToBusStop</i>	: Thời gian di chuyển từ căn hộ đến trạm xe buýt (categorical).
<i>TimeToSubway</i>	: Thời gian di chuyển từ căn hộ đến ga tàu điện ngầm (categorical).
<i>N_APT</i>	: Số căn hộ trong khu chung cư.
<i>N_manager</i>	: Số người quản lý các tiện ích tại căn hộ (ví dụ như an ninh, vệ sinh,...).
<i>N_elevators</i>	: Tổng số thang máy trong một khu chung cư.
<i>SubwayStation</i>	: Tên của ga tàu điện ngầm gần căn hộ.
<i>N_FacilitiesNearBy(PublicOffice)</i>	: Số lượng văn phòng công cộng gần căn hộ.
<i>N_FacilitiesNearBy(Hospital)</i>	: Số lượng bệnh viện gần căn hộ.
<i>N_FacilitiesNearBy(Dpartmentstore)</i>	: Số lượng cửa hàng bách hoá gần căn hộ.
<i>N_FacilitiesNearBy(Mall)</i>	: Số lượng trung tâm mua sắm gần căn hộ.

<i>N_FacilitiesNearBy(ETC)</i>	: Số lượng khách sạn, trường học đặc biệt gần căn hộ.
<i>N_FacilitiesNearBy(Park)</i>	: Số lượng công viên gần căn hộ.
<i>N_SchoolNearBy(Elementary)</i>	: Số lượng trường tiểu học gần căn hộ.
<i>N_SchoolNearBy(Middle)</i>	: Số lượng trường trung học cơ sở gần căn hộ.
<i>N_SchoolNearBy(High)</i>	: Số lượng trường trung học phổ thông gần căn hộ.
<i>N_SchoolNearBy(University)</i>	: Số lượng trường đại học gần căn hộ.
<i>N_FacilitiesInApt</i>	: Số lượng tiện ích cho người ở chung cư (ví dụ như hồ bơi, phòng gym, sân chơi ngoài trời,...).
<i>N_FacilitiesNearBy(Total)</i>	: Tổng số tiện ích gần căn hộ.
<i>N_SchoolNearBy(Total)</i>	: Tổng số trường học gần căn hộ.

2 Tiền xử lý dữ liệu

2.1 Load dữ liệu

Ta load dữ liệu từ file: *7_Daegu_Real_Estate_data.csv*

	SalePrice	YearBuilt	YrSold	MonthSold	Size(sqf)	Floor	HallwayType	HeatingType	AptManageType	N_Parkinglot(Ground)	...
0	141592	2006	2007	8	814	3	terraced	individual_heating	management_in_trust	111.0	...
1	51327	1985	2007	8	587	8	corridor	individual_heating	self_management	80.0	...
2	48672	1985	2007	8	587	6	corridor	individual_heating	self_management	80.0	...
3	380530	2006	2007	8	2056	8	terraced	individual_heating	management_in_trust	249.0	...
4	221238	1993	2007	8	1761	3	mixed	individual_heating	management_in_trust	523.0	...

5 rows × 30 columns

2.2 Một số thông tin cơ bản của dữ liệu

2.2.1 In ra thông tin của dữ liệu

Ta sử dụng hàm *DataFrame.info()* để xuất ra những thông tin của dữ liệu.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5891 entries, 0 to 5890
Data columns (total 30 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   SalePrice                                5891 non-null   int64
1   YearBuilt                                5891 non-null   int64
2   YrSold                                  5891 non-null   int64
3   MonthSold                               5891 non-null   int64
4   Size(sqf)                              5891 non-null   int64
5   Floor                                   5891 non-null   int64
6   HallwayType                             5891 non-null   object
7   HeatingType                             5891 non-null   object
8   AptManageType                           5891 non-null   object
9   N_Parkinglot(Ground)                    5891 non-null   float64
10  N_Parkinglot(Basement)                   5891 non-null   float64
11  TimeToBusStop                            5891 non-null   object
12  TimeToSubway                             5891 non-null   object
13  N_APT                                    5891 non-null   float64
14  N_manager                               5891 non-null   float64
15  N_elevators                             5891 non-null   float64
16  SubwayStation                           5891 non-null   object
17  N_FacilitiesNearBy(PublicOffice)         5891 non-null   float64
18  N_FacilitiesNearBy(Hospital)             5891 non-null   int64
19  N_FacilitiesNearBy(Dpartmentstore)       5891 non-null   float64
20  N_FacilitiesNearBy(Mall)                 5891 non-null   float64
21  N_FacilitiesNearBy(ETC)                 5891 non-null   float64
22  N_FacilitiesNearBy(Park)                 5891 non-null   float64
23  N_SchoolNearBy(Elementary)               5891 non-null   float64
24  N_SchoolNearBy(Middle)                   5891 non-null   float64
25  N_SchoolNearBy(High)                     5891 non-null   float64
26  N_SchoolNearBy(University)               5891 non-null   float64
27  N_FacilitiesInApt                       5891 non-null   int64
28  N_FacilitiesNearBy(Total)                5891 non-null   float64
29  N_SchoolNearBy(Total)                   5891 non-null   float64
dtypes: float64(16), int64(8), object(6)
memory usage: 1.3+ MB

```

Từ thông tin trên, ta biết được dữ liệu có 30 cột và 5891 dòng. Tiếp theo ta sẽ thống kê số lượng features theo từng loại dữ liệu.

```

float64    16
int64       8
object      6
dtype: int64

```

Dữ liệu của ta gồm có 16 cột dữ liệu loại *float64*, 8 cột loại *int64* và 6 cột loại *object*. Mặt khác, ta cũng có thể nói dữ liệu gồm có 24 cột dữ liệu dạng số (numerical feature) và 6 cột dạng object (object feature).

2.2.2 In ra mô tả về dữ liệu

Ta in ra mô tả về các cột dữ liệu dạng số của dữ liệu gốc.

	count	mean	std	min	25%	50%	75%	max
SalePrice	5891.0	221218.112545	106384.186446	32743.0	144247.0	207964.0	291150.0	585840.0
YearBuilt	5891.0	2002.967408	8.811782	1978.0	1993.0	2006.0	2008.0	2015.0
YrSold	5891.0	2012.691563	2.905488	2007.0	2010.0	2013.0	2015.0	2017.0
MonthSold	5891.0	6.160244	3.387752	1.0	3.0	6.0	9.0	12.0
Size(sqf)	5891.0	955.569173	382.464050	135.0	644.0	910.0	1149.0	2337.0
Floor	5891.0	12.026311	7.548743	1.0	6.0	11.0	17.0	43.0
N_Parkinglot(Ground)	5891.0	195.883551	218.597210	0.0	11.0	100.0	249.0	713.0
N_Parkinglot(Basement)	5891.0	570.761670	408.621075	0.0	184.0	536.0	798.0	1321.0
N_APT	5891.0	5.613648	2.811831	1.0	3.0	7.0	8.0	13.0
N_manager	5891.0	6.310304	3.174088	1.0	5.0	6.0	8.0	14.0
N_elevators	5891.0	11.145985	7.806232	0.0	5.0	11.0	16.0	27.0
N_FacilitiesNearBy(PublicOffice)	5891.0	4.141572	1.793496	0.0	3.0	5.0	5.0	7.0
N_FacilitiesNearBy(Hospital)	5891.0	1.296045	0.479760	0.0	1.0	1.0	2.0	2.0
N_FacilitiesNearBy(Dpartmentstore)	5891.0	0.896282	0.810964	0.0	0.0	1.0	2.0	2.0
N_FacilitiesNearBy(Mall)	5891.0	0.941436	0.401355	0.0	1.0	1.0	1.0	2.0
N_FacilitiesNearBy(ETC)	5891.0	1.941266	2.201392	0.0	0.0	1.0	5.0	5.0
N_FacilitiesNearBy(Park)	5891.0	0.654218	0.658320	0.0	0.0	1.0	1.0	2.0
N_SchoolNearBy(Elementary)	5891.0	3.022407	0.954575	0.0	2.0	3.0	4.0	6.0
N_SchoolNearBy(Middle)	5891.0	2.417756	1.037898	0.0	2.0	3.0	3.0	4.0
N_SchoolNearBy(High)	5891.0	2.659311	1.556041	0.0	1.0	2.0	4.0	5.0
N_SchoolNearBy(University)	5891.0	2.764726	1.489289	0.0	2.0	2.0	4.0	5.0
N_FacilitiesInApt	5891.0	5.809540	2.330804	1.0	4.0	5.0	7.0	10.0
N_FacilitiesNearBy(Total)	5891.0	9.870820	3.450319	0.0	8.0	9.0	13.0	16.0
N_SchoolNearBy(Total)	5891.0	10.864200	4.438513	0.0	7.0	10.0	15.0	17.0

Từ mô tả trên, ta có thể nhìn thấy một số cột chứa giá trị ngoại lai như *SalePrice*, *Size(sqf)*, *N_Parkinglot(Ground)*, *N_Parkinglot(Basement)*. Hơn nữa, cột *N_Parkinglot(Ground)* bị lệch khá nhiều, trong khi đó các cột khác bị lệch không đáng kể. Ta tiếp tục in ra mô tả về các cột có dạng object.

	count	unique	top	freq
HallwayType	5891	3	terraced	3564
HeatingType	5891	2	individual_heating	5591
AptManageType	5891	2	management_in_trust	5542
TimeToBusStop	5891	3	0~5min	4509
TimeToSubway	5891	5	0-5min	2759
SubwayStation	5891	8	Kyungbuk_uni_hospital	1644

Mô tả trên cho ta thấy được số lượng giá trị rời rạc trong từng cột, số lượng của những giá trị có tần suất xuất hiện nhiều nhất.

2.2.3 Missing values

Ta thống kê số lượng missing values có trong dữ liệu.

	Missing values	% Missing
SalePrice	0	0.0
YearBuilt	0	0.0
YrSold	0	0.0
MonthSold	0	0.0
Size(sqf)	0	0.0
Floor	0	0.0
HallwayType	0	0.0
HeatingType	0	0.0
AptManageType	0	0.0
N_Parkinglot(Ground)	0	0.0
N_Parkinglot(Basement)	0	0.0
TimeToBusStop	0	0.0
TimeToSubway	0	0.0
N_APT	0	0.0
N_manager	0	0.0
N_elevators	0	0.0
SubwayStation	0	0.0
N_FacilitiesNearBy(PublicOffice)	0	0.0
N_FacilitiesNearBy(Hospital)	0	0.0
N_FacilitiesNearBy(Dpartmentstore)	0	0.0
N_FacilitiesNearBy(Mall)	0	0.0
N_FacilitiesNearBy(ETC)	0	0.0
N_FacilitiesNearBy(Park)	0	0.0
N_SchoolNearBy(Elementary)	0	0.0
N_SchoolNearBy(Middle)	0	0.0
N_SchoolNearBy(High)	0	0.0
N_SchoolNearBy(University)	0	0.0
N_FacilitiesInApt	0	0.0
N_FacilitiesNearBy(Total)	0	0.0
N_SchoolNearBy(Total)	0	0.0

Dữ liệu của ta không có cột nào bị mất dữ liệu.

3 Phân tích dữ liệu

Vì dữ liệu có số lượng features khá nhiều nên ta cần chọn ra những feature quan trọng từ dữ liệu để có thể trực quan một cách dễ dàng hơn.

Để có thể chọn ra những feature quan trọng trong bộ dữ liệu này, ta cần tìm hiểu những nhu cầu thiết yếu mà khách hàng cần ở một căn hộ. Một cách trực quan, ta có thể đưa ra những suy luận về một số nhân tố ảnh hưởng đến hành vi mua hàng của khách hàng như sau

Thiết kế : Mua căn hộ để ở, người mua rất muốn sở hữu một không gian không chỉ phù hợp

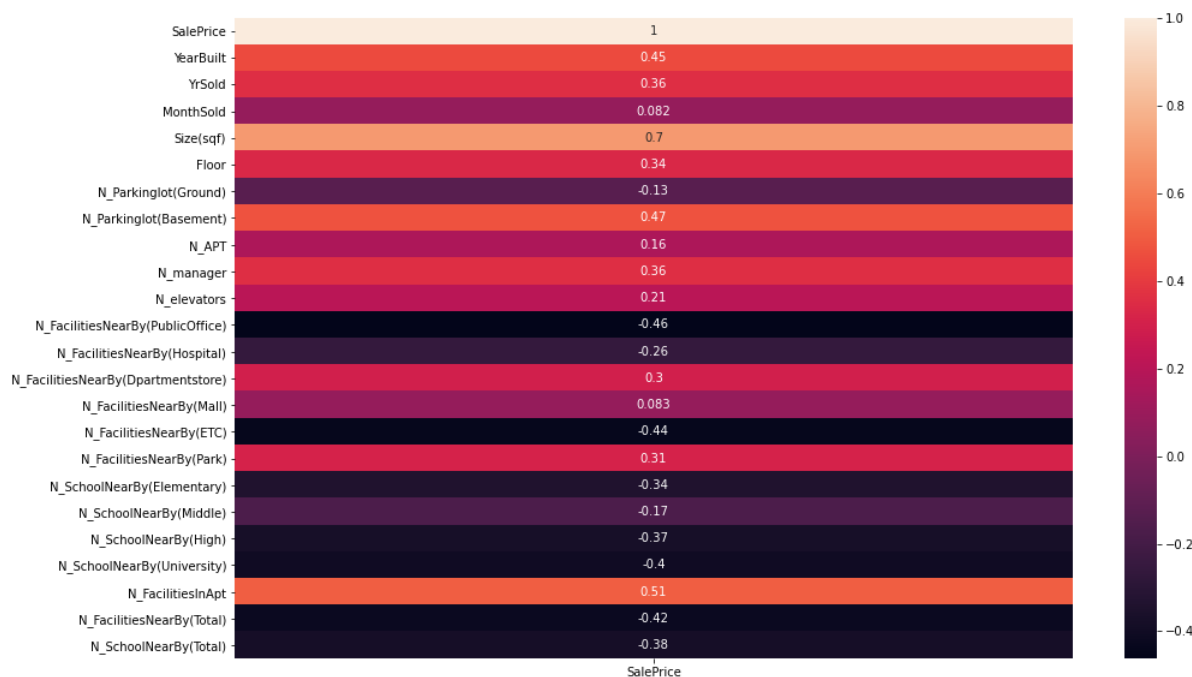
về kích thước mà còn phải có sự tiện nghi cho cuộc sống của mình. Cụ thể hơn, người mua quan tâm đến diện tích và những cơ sở vật chất bên trong mà căn hộ mang lại.

Tuổi thọ: Tuổi của một căn hộ là một trong những yếu tố quan trọng mà người mua sẽ quan tâm vì nó có thể ảnh hưởng đến giá thành cũng như chất lượng cuộc sống của họ tại đây.

Vị trí căn hộ: Bên cạnh việc chọn một căn hộ có thiết kế hợp lý, người mua cũng mong muốn sở hữu một căn hộ có vị trí thuận lợi cho việc di chuyển của họ. Cụ thể hơn, người mua sẽ ưu tiên chọn những căn hộ gần với các những phương tiện công cộng để thuận tiện trong việc di chuyển và tiết kiệm thời gian, đặc biệt là những khách hàng ở độ tuổi đi làm.

Với những suy luận trên, ta chọn ra những features có tầm ảnh hưởng lớn đến dữ liệu của mình gồm *SalePrice*, *YearBuilt*, *YrSold*, *Size(sqf)*, *TimeToBusStop*, *TimeToSubway*, *N_FacilitiesInApt*.

Ta vẽ ra *heatmap* để mô tả sự tương quan giữa các numerical features với target feature (*SalePrice*).

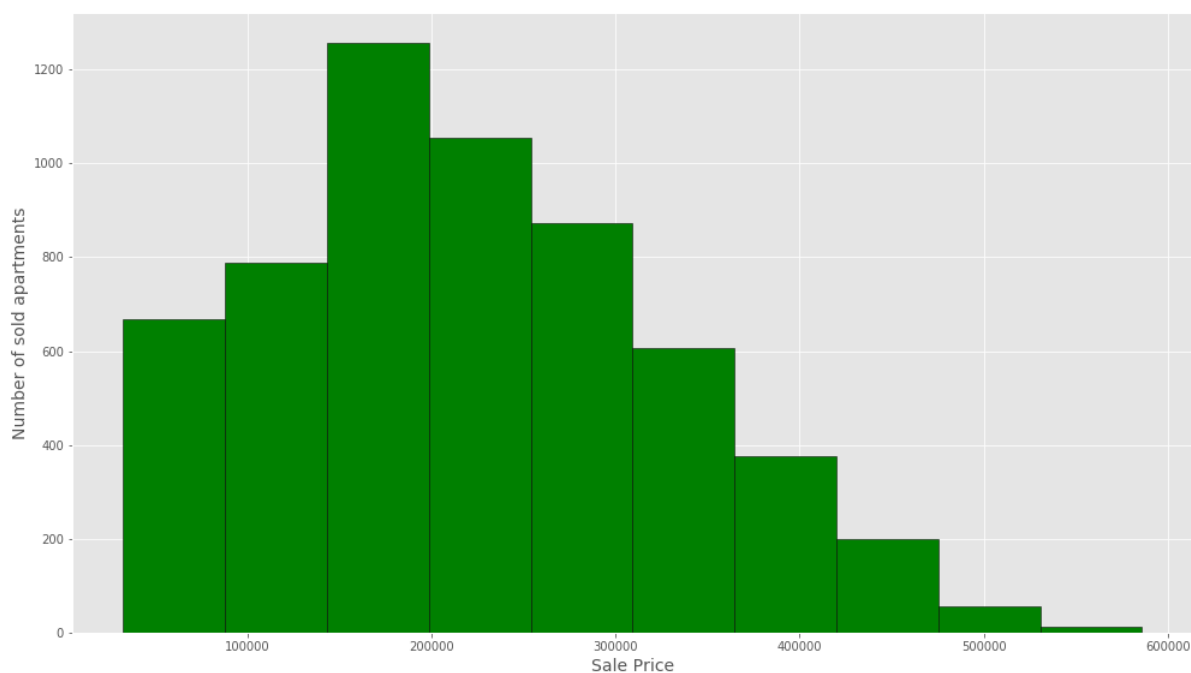


Từ heatmap ở trên, ta thấy được hệ số tương quan giữa các features *YearBuilt*, *YrSold*, *Size(sqf)*, *N_FacilitiesInApt* với target feature (*SalePrice*) là 0.45, 0.36, 0.7, 0.51. Các hệ số này đều khá cao nên những features ta đã chọn cũng khá hợp lý.

4 Trực quan hoá dữ liệu

4.1 Thống kê và vẽ các biểu đồ biểu diễn các features quan trọng

4.1.1 SalePrice



Nhận xét:

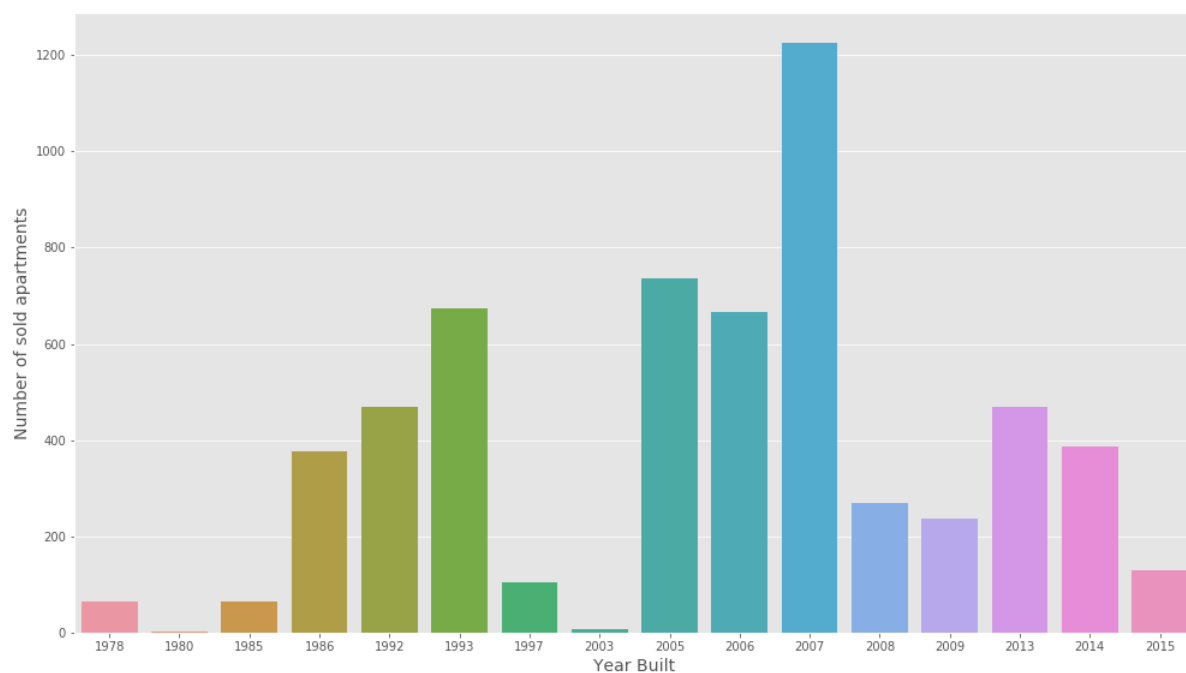
- Ta thấy đồ thị có dạng gần với phân phối chuẩn và bị lệch phải.
- Đa số các căn hộ được bán với giá trong khoảng từ 150000 - 200000 (US dollar).
- Các căn hộ có giá bán từ 200000 US dollar trở thì bán được càng ít.

4.1.2 YearBuilt

Thống kê số lượng căn hộ giao dịch theo từng năm xây dựng.

	YearBuilt	Number of Apt
0	1978	65
1	1980	4
2	1985	65
3	1986	378
4	1992	469
5	1993	674
6	1997	106
7	2003	9
8	2005	737
9	2006	665
10	2007	1225
11	2008	270
12	2009	238
13	2013	468
14	2014	387

Trực quan bằng đồ thị.

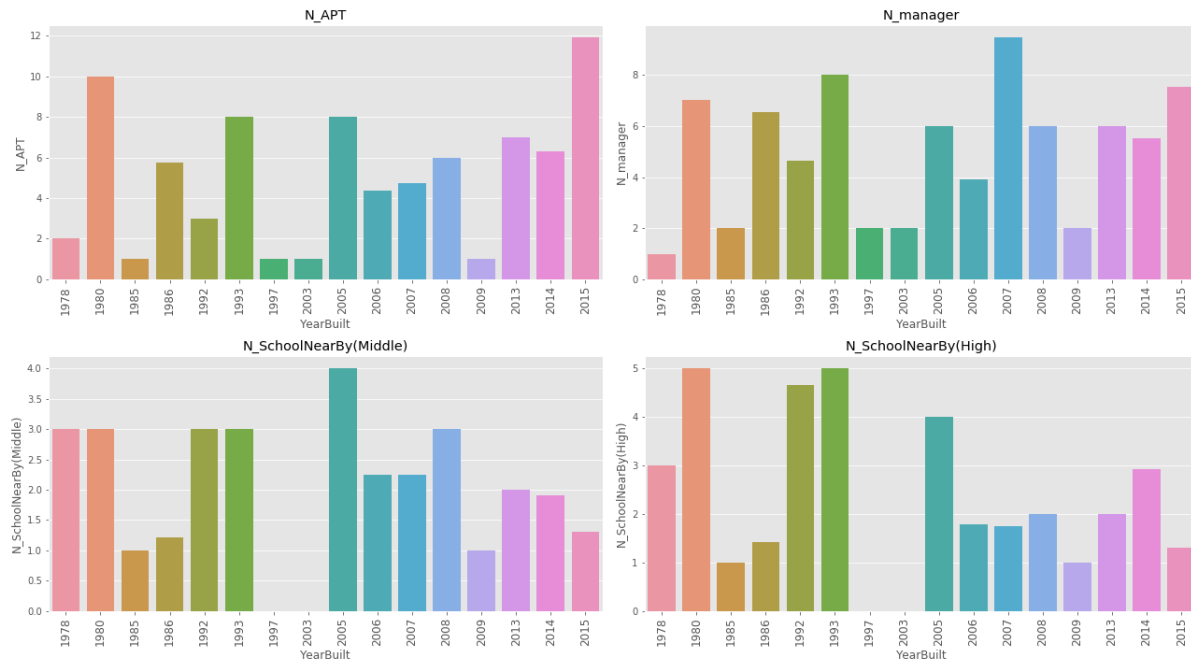


Từ đồ thị, ta thấy

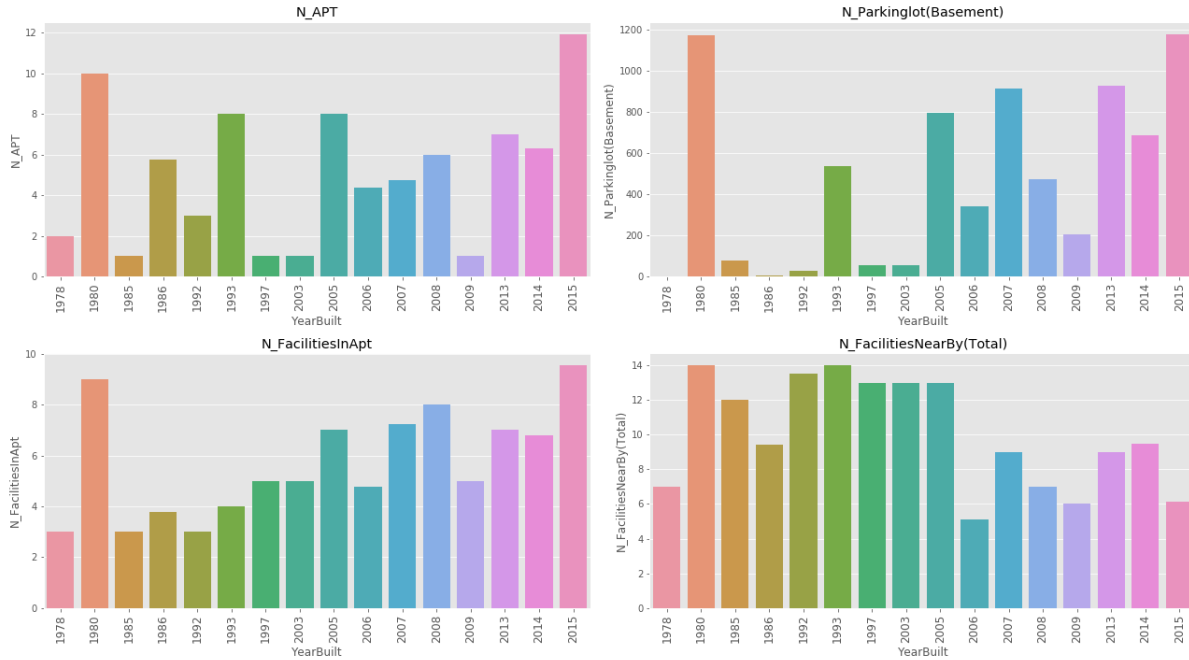
- Số lượng căn hộ bán ra được xây dựng trong giai đoạn từ 1978 - 1985 rất ít, đặc biệt chỉ có 4 căn hộ được bán ra xây dựng trong năm 1980.
- Số lượng căn hộ bán ra được xây dựng trong giai đoạn từ 1997 - 2003 rất ít, đặc biệt chỉ có 9 căn hộ được bán ra xây dựng trong năm 2003.

- Số lượng các căn hộ bán ra được xây dựng từ năm 2005 trở đi tăng đột ngột, trong đó các căn hộ được xây dựng vào năm 2007 được bán ra nhiều nhất với số lượng 1225 căn hộ.

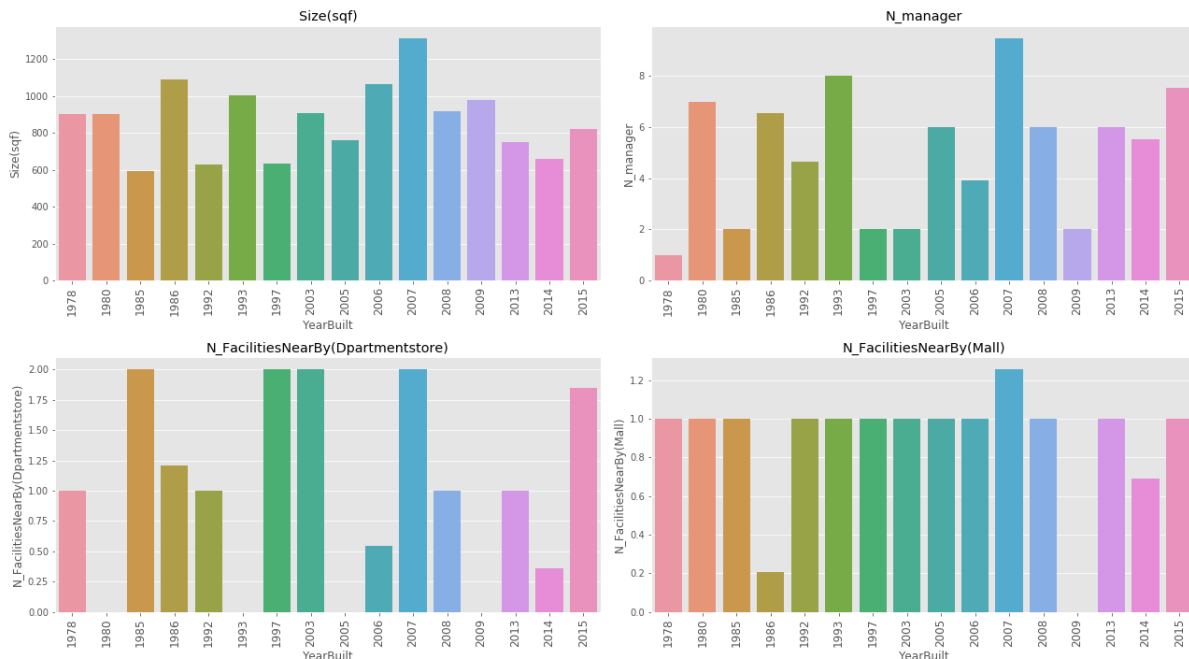
Để giải thích cho những vấn đề trên, ta sẽ vẽ thêm một số biểu đồ để có cái nhìn bao quát hơn.



Với 4 biểu đồ trên, ta tập trung vào giai đoạn 1997 - 2003. Ta thấy rằng các căn hộ được xây dựng trong giai đoạn này hầu như không gần các trường cấp 2, cấp 3, khả năng quản lý kém (số lượng người làm an ninh, vệ sinh,... ít) và xung quanh chỉ có 1 căn hộ láng giềng. Do đó ta có thể kết luận rằng tính an toàn của những căn hộ này khá thấp nên khả năng được chọn mua không cao.



Với 4 biểu đồ trên, ta tập trung vào năm 1980. Điều đặc biệt ở các căn hộ giao dịch được xây dựng vào năm 1980 là số lượng giao dịch đặc biệt thấp. Tuy nhiên, theo những đồ thị trên thì những căn hộ này có rất nhiều làng giềng, số lượng chỗ để xe ở tầng hầm nhiều, có nhiều tiện ích ở cả trong lẫn ngoài căn hộ. Điều này có thể suy ra được, những căn hộ giao dịch được xây dựng trong năm 1980 là những căn hộ cũ có chất lượng tốt và giá rẻ nhưng lại hiếm nên số lượng bán ra khá thấp.



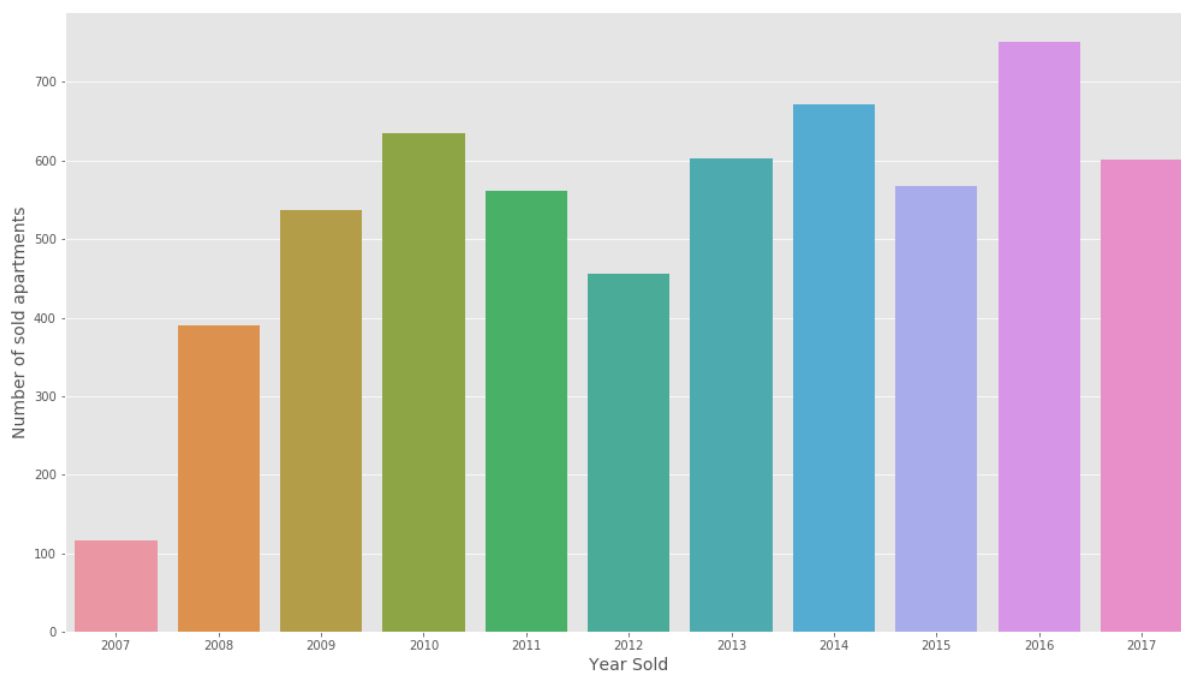
Với 4 biểu đồ trên, ta tập trung vào năm 2007. Những căn hộ giao dịch được xây dựng trong năm 2007 đặc biệt có kích thước trung bình lớn nhất (khoảng hơn 1200 square feet), số lượng người quản lý và làm dịch vụ cao (hơn 8 người), vị trí lại gần những trung tâm thương mại, các dịch vụ tiện ích công cộng. Điều này khiến cho số lượng căn hộ xây dựng trong năm này được chú ý và mua lại nhiều nhất.

4.1.3 YrSold

Thống kê số lượng căn hộ giao dịch tính theo năm giao dịch.

	YrSold	Number of Apt
0	2007	117
1	2008	390
2	2009	537
3	2010	634
4	2011	562
5	2012	456
6	2013	603
7	2014	672
8	2015	568
9	2016	751
10	2017	601

Trực quan bằng đồ thị.



Từ đồ thị, ta thấy:

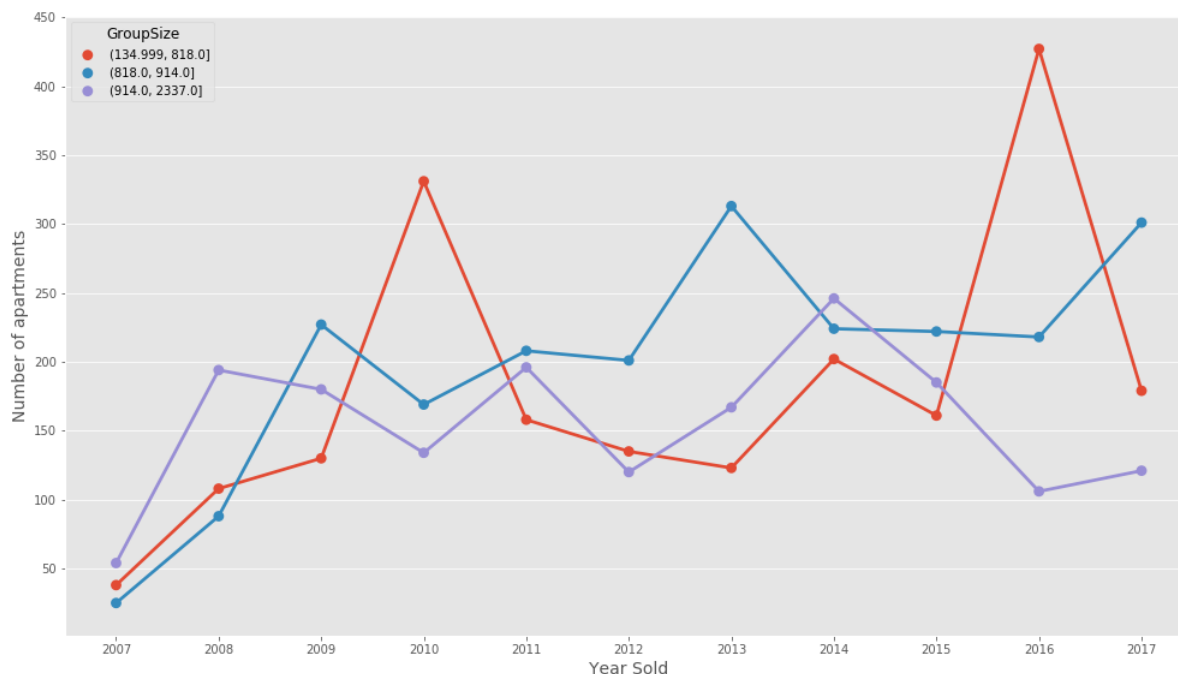
- Năm 2007, số lượng căn hộ được bán rất thấp chỉ 117 căn hộ. Lý do là vì dữ liệu bắt đầu lấy từ tháng 8/2007 nên số lượng căn hộ bán ra ở đây chỉ tính cho 4 tháng cuối của năm 2007.
- Nếu không tính năm 2007 thì năm 2008 sẽ có số lượng căn hộ được bán ra thấp nhất với 390 căn hộ.
- Năm 2016 có số lượng căn hộ bán ra nhiều nhất với 751 căn hộ.

4.1.4 Size(sqf)

Thống kê số lượng căn hộ giao dịch theo kích thước và năm giao dịch.

	YrSold	GroupSize	Number of Apt
0	2007	(134.999, 818.0]	38
1	2007	(818.0, 914.0]	25
2	2007	(914.0, 2337.0]	54
3	2008	(134.999, 818.0]	108
4	2008	(818.0, 914.0]	88

Trực quan bằng đồ thị



Từ đồ thị trên, ta thấy rằng:

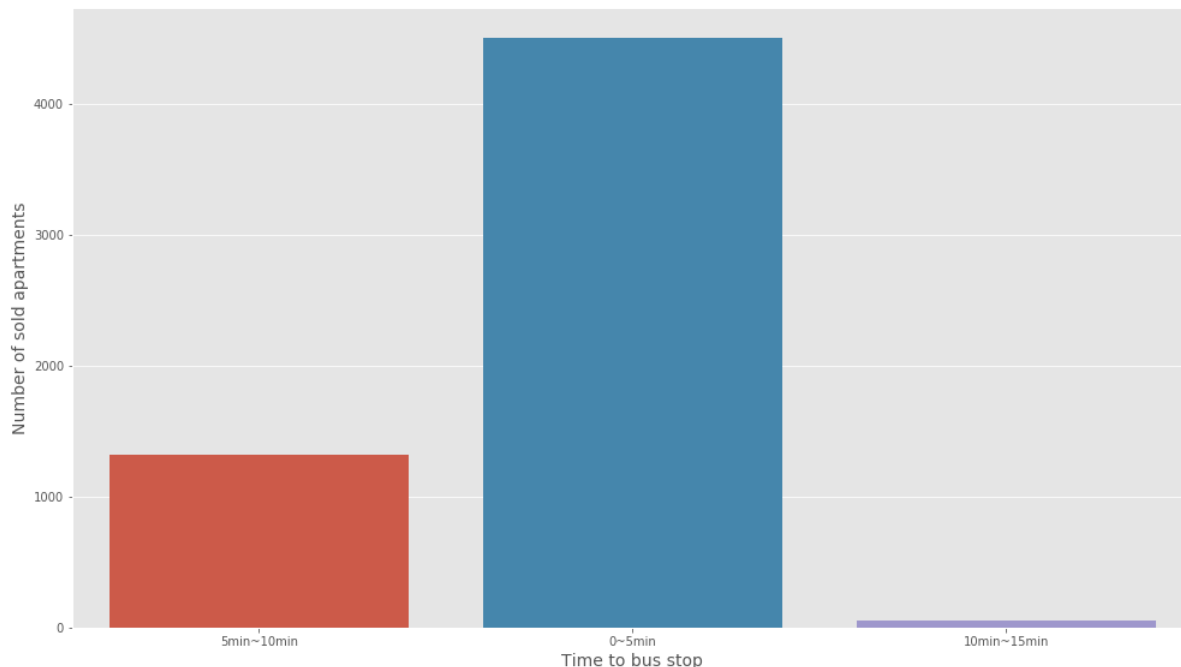
- Nhìn chung ta có thể thấy số giao dịch của mỗi loại kích thước căn hộ đều có biến động, nhưng các căn hộ có kích thước nhỏ lại có số lượng biến động lớn nhất. Số lượng giao dịch cũng tăng đáng kể ở năm 2017 so với 2007, 2008.
- Số lượng giao dịch của các căn hộ có kích thước lớn mặc dù có độ biến động từ 2008 - 2016 nhưng nhìn chung vẫn có số lượng khá thấp so với 2 loại kích thước còn lại.
- Trong giai đoạn từ 2013 - 2015, số lượng giao dịch của các căn hộ có kích thước nhỏ lại thấp hơn những căn hộ có kích thước lớn. Điều này cho thấy trong giai đoạn này, người dân có xu hướng chọn mua những căn hộ có kích thước trung bình hoặc lớn nhiều hơn.
- Trong khoảng thời gian năm 2009 tới 2010, số lượng căn hộ có kích thước nhỏ được bán ra trong thời gian này tăng đột biến (gấp gần 3 lần) trong khi số lượng căn hộ bán ra có kích thước vừa và lớn lại sụt giảm đáng kể.
- Chúng ta cũng có xu hướng tương tự vào giai đoạn năm 2015, 2016 khi số lượng căn hộ có kích thước nhỏ bán ra trong thời gian này cũng tăng đột biến (khoảng gấp 3 lần) trong khi số lượng căn hộ có kích thước lớn bán ra lại sụt giảm khoảng 1 nửa trong khi số lượng có kích thước vừa giữ ổn định.

4.1.5 TimeToBusStop

Thống kê số lượng căn hộ giao dịch theo thời gian đến các trạm xe buýt.

	TimeToBusStop	Number of Apt
0	0~5min	4509
1	10min~15min	55
2	5min~10min	1327

Trực quan bằng đồ thị.



Từ đồ thị trên, ta thấy rằng:

- Đa số các căn hộ bán ra đều cách trạm xe buýt từ 0 - 5 phút đi bộ.
- Các căn hộ ở xa trạm xe buýt hơn thường không được lựa chọn để mua nên có số lượng cực thấp.

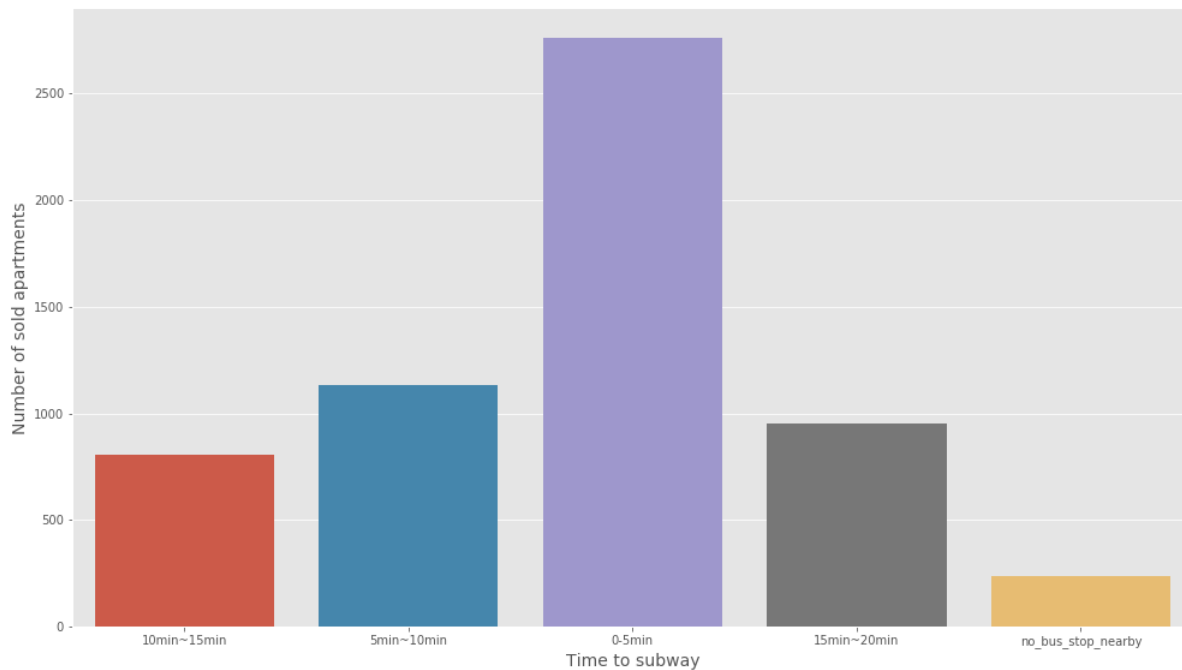
Từ 2 điều trên ta có thể suy ra rằng người dân tại thành phố Daegu ưa chuộng những căn hộ gần với trạm xe buýt hơn vì điều đó sẽ khiến họ trở nên thuận tiện trong việc di chuyển.

4.1.6 TimeToSubway

Thống kê số lượng căn hộ giao dịch theo thời gian đến các trạm tàu điện ngầm.

	TimeToSubway	Number of Apt
0	0-5min	2759
1	10min~15min	806
2	15min~20min	953
3	5min~10min	1135
4	no_bus_stop_nearby	238

Trực quan bằng đồ thị.



Từ đồ thị trên, ta có thể thấy:

- Đa số các căn hộ cách trạm tàu điện ngầm từ 0 - 5 phút đi bộ được chọn mua nhiều nhất.
- Các căn hộ không gần với trạm xe nào được chọn mua ít nhất.

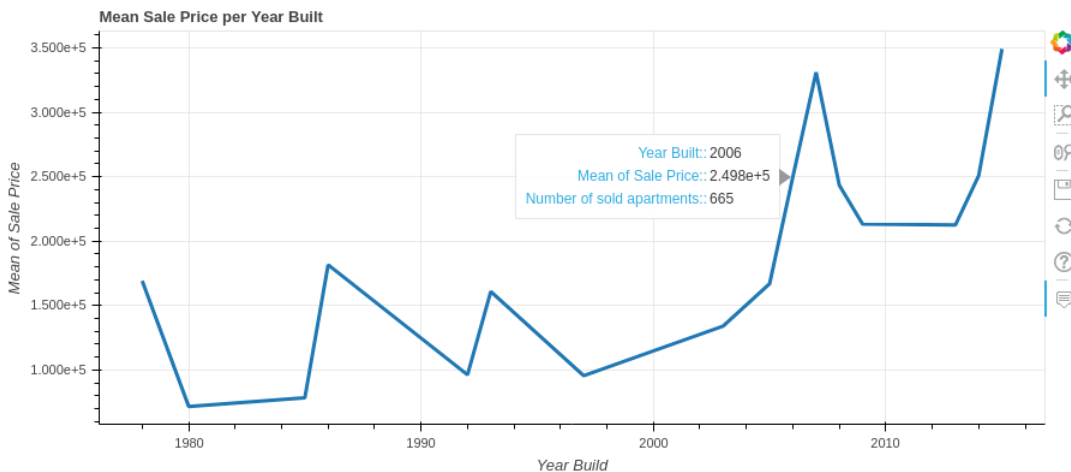
Những điều này cho thấy rằng dân cư ở thành phố Daegu ưu tiên lựa chọn những căn hộ gần các trạm tàu điện ngầm để thuận tiện cho việc di chuyển của họ. Những căn hộ không gần các trạm xe là lựa chọn cuối cùng của họ nếu không tìm được những căn hộ tốt hơn nên số lượng căn hộ bán ra rất ít.

4.2 Thống kê và vẽ các biểu đồ biểu diễn mối quan hệ giữa các features

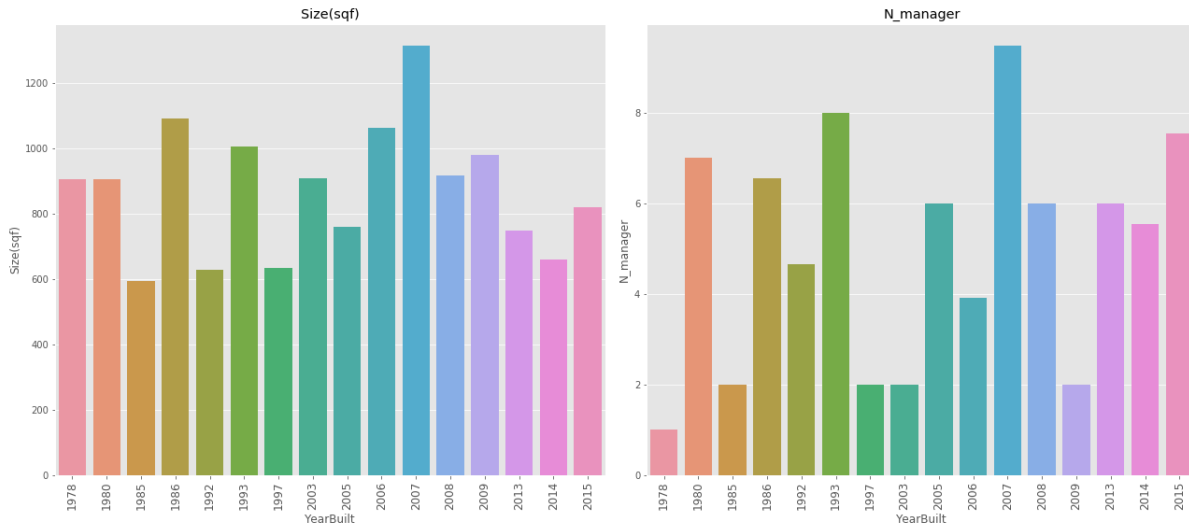
4.2.1 Thống kê số lượng và giá bán trung bình của các căn hộ theo từng năm xây dựng

	YearBuilt	MeanSalePrice	NoSoldApt
0	1978	169039.615385	65
1	1980	71681.000000	4
2	1985	78401.092308	65
3	1986	181543.283069	378
4	1992	96215.933902	469
5	1993	160855.087537	674
6	1997	95416.122642	106
7	2003	134021.111111	9
8	2005	166852.907734	737
9	2006	249809.592481	665
10	2007	330971.122449	1225
11	2008	243339.081481	270
12	2009	213169.630252	238
13	2013	212606.918803	468
14	2014	250990.095607	387
15	2015	349141.549618	131

Trực quan bằng đồ thị.



(*Interactive Visualization*) Một số thao tác có thể thực hiện ở đồ thị trên: Một số thông tin về năm xây dựng, giá giao dịch trung bình, số lượng giao dịch sẽ được hiện ra khi ta đưa con trỏ vào từng mốc thời gian trên đồ thị, ngoài ra ta cũng có thể di chuyển, phóng to, thu nhỏ đồ thị để tiện cho việc quan sát.



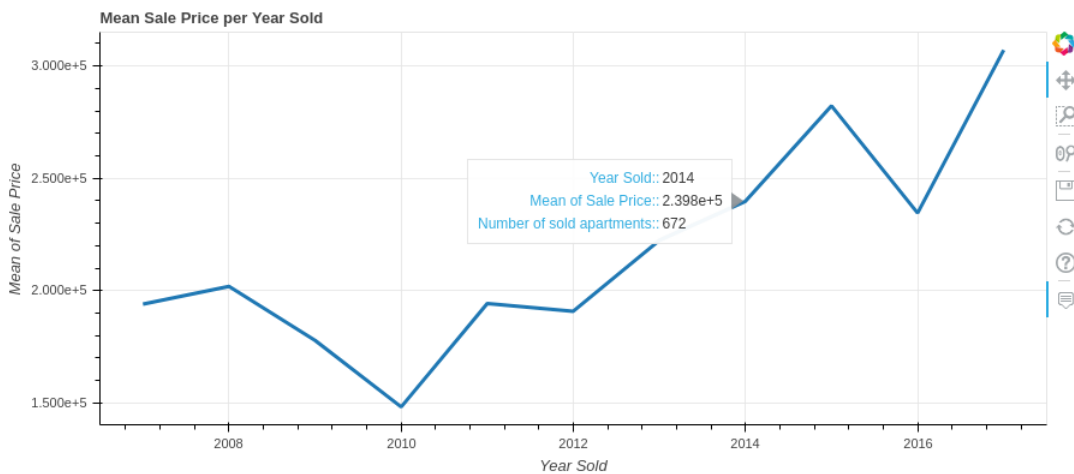
Từ những đồ thị trên, ta thấy rằng:

- Trong giai đoạn từ 1978 - 1997, giá giao dịch trung bình của các căn hộ được xây dựng trong giai đoạn này có nhiều biến động. Trong đó năm 1980 có giá giao dịch trung bình thấp nhất, điều này có thể là do số lượng các căn hộ giao dịch được xây dựng trong năm này rất thấp (chỉ 4 căn hộ) và có tuổi thọ cao.
- Giai đoạn sau năm 1997 đến năm 2007, giá giao dịch trung bình của các căn hộ được xây dựng trong giai đoạn này tăng đột biến. Lý giải cho việc này là những căn hộ giao dịch được xây dựng trong giai đoạn này có xu hướng tăng về kích thước và số lượng người quản lí.
- Tuy nhiên, các căn hộ được xây dựng từ năm 2007 đến 2013 lại có dấu hiệu giảm. Lý giải cho điều này có thể là do số lượng giao dịch các căn hộ được xây dựng trong giai đoạn này giảm đi và đa số các căn hộ giao dịch đều có kích thước nhỏ và trung bình nên giá giao dịch trung bình của các căn hộ được xây dựng trong giai đoạn này giảm so với trước đó.
- Từ năm 2013 trở đi, giá giao dịch trung bình của các căn hộ được xây dựng trong giai đoạn này bắt đầu tăng trở lại. Lý do vì người dân bắt đầu chọn mua những căn hộ có kích thước trung bình và lớn nhiều hơn, do có kích thước lớn hơn nên giá cũng tăng không ít.

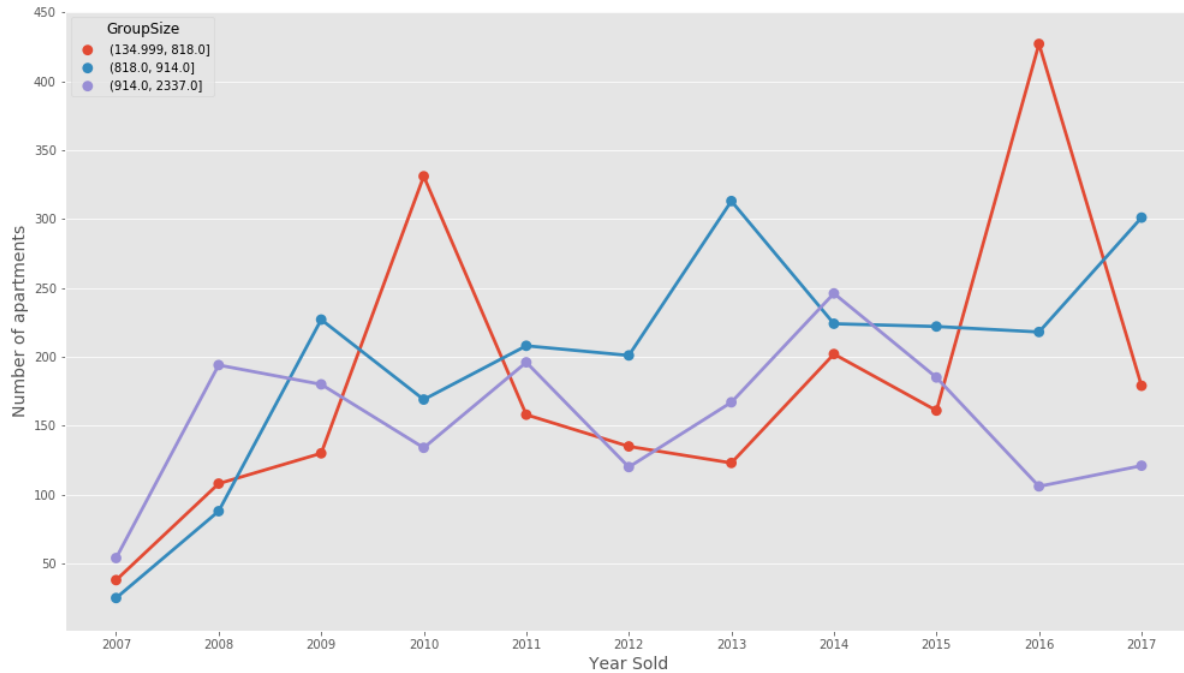
4.2.2 Thống kê số lượng và giá bán trung bình của các căn hộ theo từng năm bán ra

	YrSold	MeanSalePrice	NoSoldApt
0	2007	193989.034188	117
1	2008	201853.058974	390
2	2009	177825.657356	537
3	2010	148217.151420	634
4	2011	194249.209964	562
5	2012	190801.328947	456
6	2013	222325.983416	603
7	2014	239752.703869	672
8	2015	282221.441901	568
9	2016	234509.402130	751
10	2017	307065.311148	601

Trực quan bằng đồ thị.



(*Interactive Visualization*) Một số thao tác có thể thực hiện ở đồ thị trên: Một số thông tin về năm giao dịch, giá giao dịch trung bình, số lượng giao dịch sẽ được hiện ra khi ta đưa con trỏ vào từng mốc thời gian trên đồ thị, ngoài ra ta cũng có thể di chuyển, phóng to, thu nhỏ đồ thị để tiện cho việc quan sát.



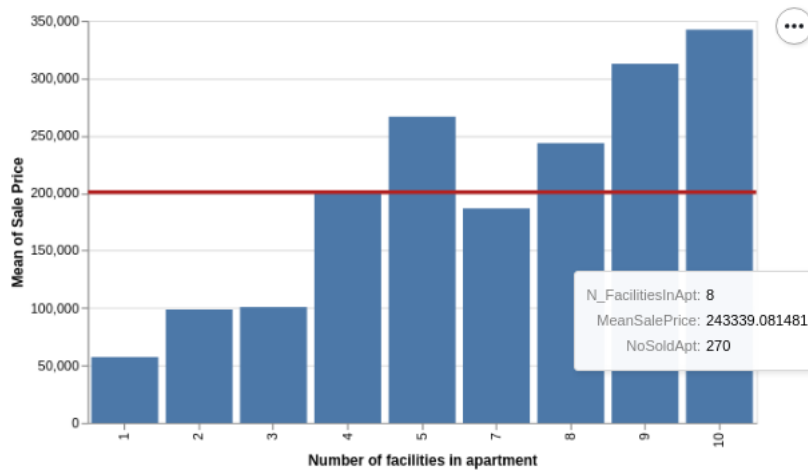
Từ đồ thị trên, ta thấy rằng:

- Nhìn chung, giá giao dịch trung bình của các căn hộ từ năm 2007 - 2017 có xu hướng tăng.
- Trong năm 2010, giá giao dịch trung bình lại đạt mức thấp nhất trong giai đoạn khảo sát. Lý giải cho điều này là vì số lượng lớn các căn hộ giao dịch trong giai đoạn này đều có kích thước nhỏ, các căn hộ có kích thước vừa và lớn lại chiếm rất ít, điều này cũng có thể giải thích tại sao ở năm 2010 số lượng giao dịch cao hơn những năm lân cận lại có giá giao dịch trung bình thấp.
- Giai đoạn từ 2010 - 2015, giá giao dịch trung bình các căn hộ có xu hướng tăng nhanh. Lý giải cho việc này là sự sụt giảm trong số lượng căn hộ có kích thước nhỏ được bán ra và tăng đối với căn hộ có kích thước vừa và lớn. Hơn nữa, chất lượng nói chung các căn hộ trong giai đoạn này đều cao.
- Tuy nhiên đến năm 2016, giá giao dịch trung bình lại bị giảm. Lý do cho điều này là số lượng các căn hộ có kích thước nhỏ trong năm này tăng đột biến, khiến cho giá giao dịch trung bình bị ảnh hưởng lớn.

4.2.3 Thống kê giá giao dịch trung bình theo số lượng tiện ích có trong căn hộ.

	N_FacilitiesInApt	MeanSalePrice	NoSoldApt
0	1	57294.836364	55
1	2	98527.188406	69
2	3	100854.293944	677
3	4	200061.286109	1447
4	5	266346.104310	1160
5	7	186586.816327	1225
6	8	243339.081481	270
7	9	312737.177340	203
8	10	342371.644586	785

Trực quan bằng đồ thị.



(*Interactive Visualization*) Một số thao tác có thể thực hiện ở đồ thị trên: Một số thông tin về số lượng tiện ích, giá giao dịch trung bình và số lượng giao dịch sẽ được hiện ra khi ta đưa con trỏ vào từng cột trên đồ thị, ngoài ra khi ta kéo thả chuột chọn vùng trên đồ thị, đường màu đỏ biểu thị giá giao dịch trung bình sẽ thay đổi theo số cột đã chọn.

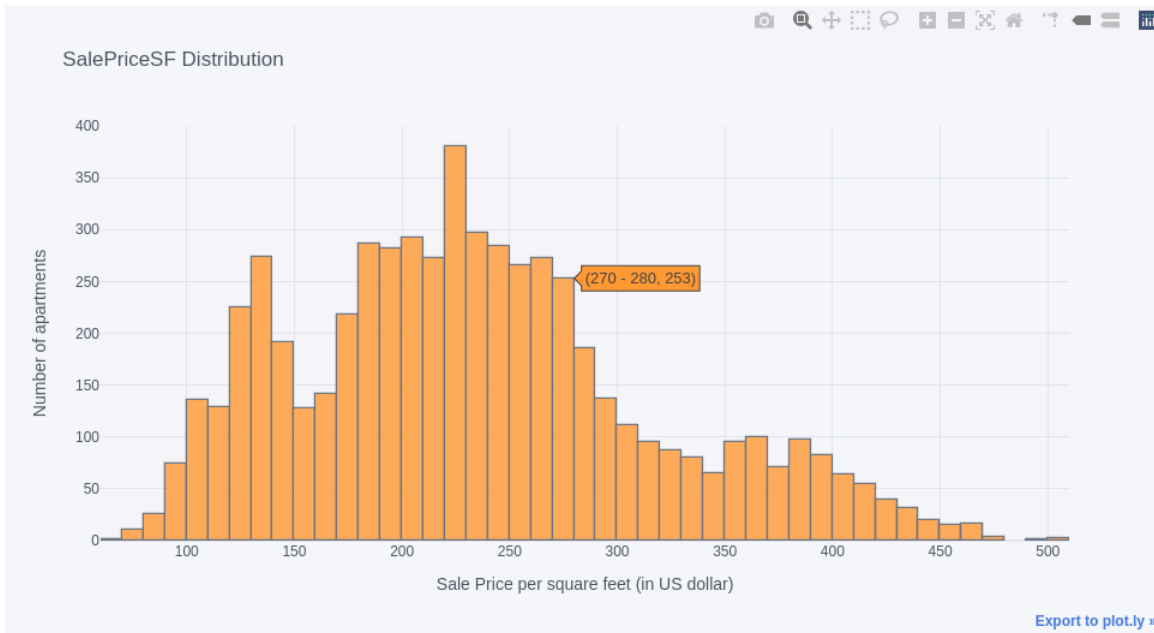
Từ đồ thị trên, ta thấy rằng:

- Khi số lượng tiện tích trong chung cư càng nhiều thì giá căn hộ bán ra càng lớn. Cụ thể là ở trên mức trung bình lượng tiền bán của tổng thể.

4.2.4 Thống kê giá bán trên một đơn vị diện tích (square feet).

	SalePrice	Size(sqf)	SalePriceSF
0	141592	814	173.945946
1	51327	587	87.439523
2	48672	587	82.916525
3	380530	2056	185.082685
4	221238	1761	125.632027

Trực quan bằng đồ thị.



(*Interactive Visualization*) Một số thao tác có thể thực hiện ở đồ thị trên: Một số thông tin về phạm vi diện tích và giá giao dịch tương ứng sẽ được hiện ra khi ta đưa con trỏ vào từng cột trên đồ thị, ngoài ra ta có thể phóng to đồ thị theo vùng được chỉ định.

Từ đồ thị trên, ta thấy rằng:

- Đa số người dân lựa chọn mua những căn hộ có giá nằm trong khoảng 220 - 230 dollars trên mỗi square feet, với số lượng giao dịch lên đến 380 căn hộ.
- Đa số người dân có thu nhập trung bình khi các căn hộ giao dịch đa số có giá trong khoảng 180 - 280 dollars trên mỗi square feet.
- Với giá giao dịch trên mỗi square feet trên 230, số lượng giao dịch càng thấp.

Từ những điều trên, ta có thể kết luận rằng, đa số người dân ở phường Daebong có mức thu nhập trung bình, những người dân có thu nhập cao vẫn có những chiếm tỉ lệ không cao.

4.2.5 Thống kê giá giao dịch trên từng square feet theo tuổi thọ căn hộ.

	SalePriceSF	ConstructionAge
0	173.945946	1
1	87.439523	22
2	82.916525	22
3	185.082685	1
4	125.632027	14

Trực quan bằng đồ thị.

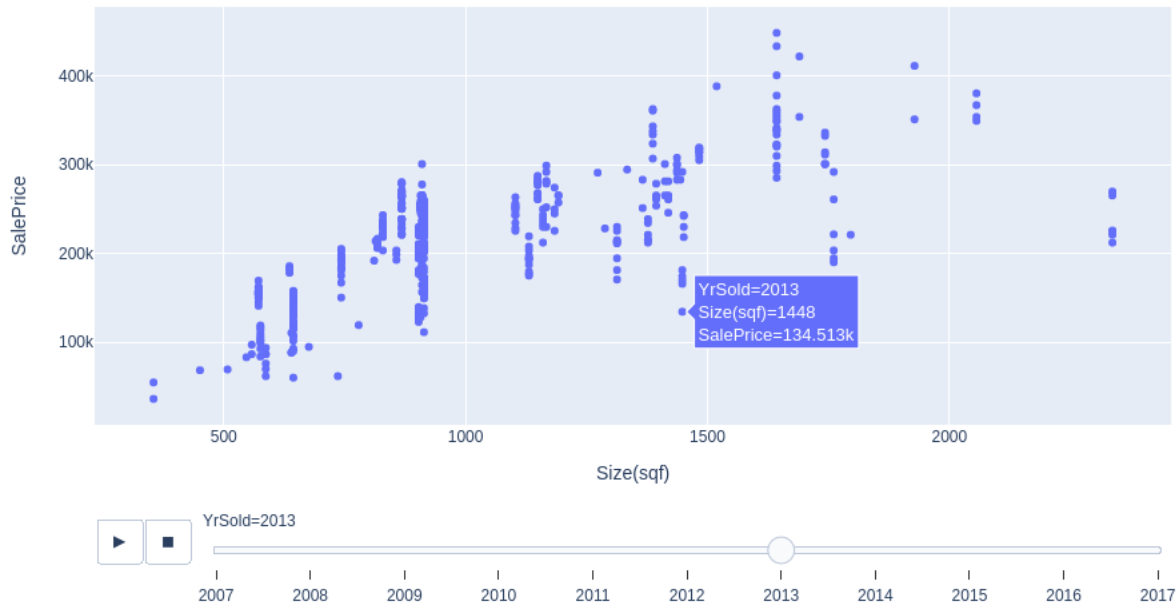


(*Interactive Visualization*) Một số thao tác có thể thực hiện ở đồ thị trên: Một số thông tin về tuổi thọ và giá trên mỗi square feet tương ứng sẽ được hiện ra khi ta đưa con trỏ vào từng điểm trên đồ thị, ngoài ra ta có thể phóng to đồ thị theo vùng được chỉ định.

Từ đồ thị trên, ta thấy rằng có 2 nhóm khách hàng chính:

- Một nhóm người chọn mua những căn hộ có tuổi thọ thấp và giá trên mỗi square feet cao
- Nhóm còn lại chọn các căn hộ có tuổi thọ cao hoặc trung bình và giá trên mỗi square feet thấp.

4.2.6 Sự thay đổi giá giao dịch theo kích thước và năm giao dịch



(*Interactive Visualization*) Một số thao tác có thể thực hiện ở đồ thị trên: Một số thông tin về năm giao dịch, kích thước và giá giao dịch tương ứng sẽ được hiện ra khi ta đưa con trỏ vào từng điểm trên đồ thị, ngoài ra đồ thị sẽ thay đổi khi ta cho thời gian (năm giao dịch) thay đổi.

Từ đồ thị trên, ta thấy rằng:

- Số lượng giao dịch ngày càng dày đặc theo thời gian.
- Số lượng giao dịch các căn hộ có kích thước lớn hơn 1000 square feet ngày càng nhiều.
- Giá thành các căn hộ giao dịch cũng tăng dần theo thời gian.

Điều này cho thấy rằng mức sống của người dân tại phường Daebong ngày càng được cải thiện do số lượng các căn hộ có kích thước lớn hơn 1000 square feet ngày càng tăng. Nhìn chung, giá tiền các căn hộ được bán ra có xu hướng tăng lên theo thời gian. Điều đó chứng tỏ cuộc sống ngày càng đắt đỏ hơn.

4.2.7 Mối quan hệ giữa tuổi thọ căn hộ và giá giao dịch trên mỗi square feet theo thời gian đến trạm xe buýt gần nhất



(*Interactive Visualization*) Một số thao tác có thể thực hiện ở đồ thị trên: Một số thông tin về tuổi thọ của căn hộ và giá giao dịch trên mỗi square feet tương ứng sẽ được hiện ra khi ta đưa con trỏ vào từng điểm trên đồ thị, ngoài ra ta cũng có thể tùy chọn đối tượng theo thời gian đến trạm xe buýt gần nhất để hiển thị trên đồ thị bằng cách nhấn vào từng mục trên chú thích.

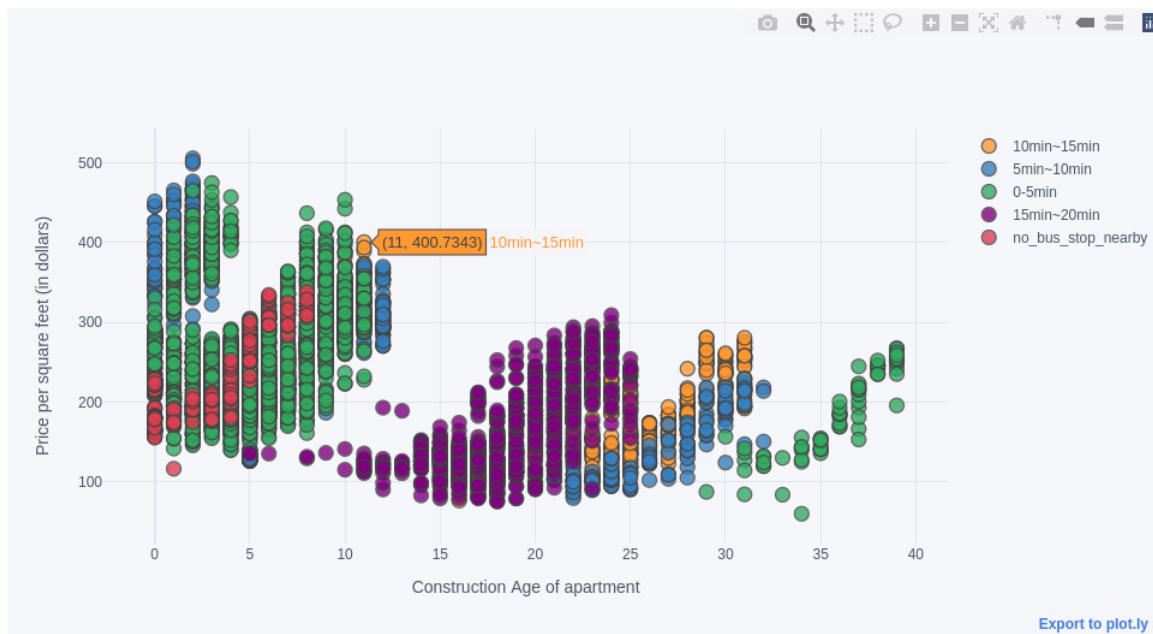
Từ đồ thị trên, ta thấy rằng:

- Đa số các căn hộ giao dịch được chọn đều là các căn hộ gần trạm xe buýt (cách trạm từ 0 - 5 phút di chuyển).
- Phần lớn các căn hộ giao dịch gần trạm xe buýt hoặc có tuổi thọ thấp và giá trên mỗi square feet cao, hoặc có tuổi thọ cao và giá trên mỗi square cao thấp.
- Các căn hộ cách xa trạm xe buýt thường là những căn hộ có tuổi thọ thấp và giá khá cao.
- Các căn hộ được giao dịch có tuổi thọ cao đều rất gần với các trạm xe buýt.

Điều này cho thấy rằng, người dân ở phường Daebong luôn ưu tiên cho việc chọn nơi ở thuận tiện cho việc di chuyển của mình bằng phương tiện công cộng là xe buýt, ta thấy được điều

này thông qua việc mặc dù những căn hộ gần trạm xe buýt có giá trên mỗi square feet cao hoặc tuổi thọ cao nhưng vẫn có số lượng căn hộ được mua vẫn chiếm đa số. Ta cũng có thể thấy nhịp sống tại đây khá tất bật khi cần phải chọn những nơi gần trạm dừng để có thể tiết kiệm thời gian.

4.2.8 Mối quan hệ giữa tuổi thọ căn hộ và giá giao dịch trên mỗi square feet theo thời gian đến trạm tàu điện ngầm gần nhất



(*Interactive Visualization*) Một số thao tác có thể thực hiện ở đồ thị trên: Một số thông tin về tuổi thọ của căn hộ và giá giao dịch trên mỗi square feet tương ứng sẽ được hiện ra khi ta đưa con trỏ vào từng điểm trên đồ thị, ngoài ra ta cũng có thể tùy chọn đối tượng theo thời gian đến trạm xe buýt gần nhất để hiển thị trên đồ thị bằng cách nhấn vào từng mục trên chú thích.

Từ đồ thị trên, ta thấy rằng:

- Đa số các căn hộ giao dịch được chọn đều là các căn hộ gần trạm tàu điện ngầm.
- Các căn hộ cách khá xa trạm tàu điện ngầm đa số tập trung vào những căn hộ có tuổi thọ trung bình (từ 10 - 25 năm).
- Các căn hộ không gần trạm dừng nào đa số đều là những căn hộ có tuổi thọ thấp (chỉ từ 0 - 8 năm).

- Các căn hộ gần với trạm tàu điện ngầm (cách 0 - 10 phút di chuyển) đều tập trung ở khu vực căn hộ có tuổi thọ thấp nhưng giá trên mỗi square feet cao hoặc tuổi thọ cao nhưng giá trên mỗi square feet thấp.
- Các căn hộ có tuổi thọ khá cao (từ 12 - 25 năm) mà cách trạm tàu điện ngầm 15 - 20 phút di chuyển thường có giá thành trên mỗi square feet thấp.

Những điều trên cho thấy rằng người dân vẫn ưu tiên chọn những căn hộ gần với trạm tàu điện ngầm hơn để thuận tiện cho việc di chuyển. Lý giải điều này thông qua việc ta có thể thấy rằng những căn hộ gần trạm dừng (cách khoảng 0 - 5 phút) đa số đều trẻ nên giá thành khá cao hoặc tuổi thọ rất cao (29 - 39 năm) những vẫn được mua với số lượng lớn. Những căn hộ cách khá xa trạm tàu điện ngầm (khoảng từ 15 - 20 phút) được chọn mua đa số là những căn hộ có giá thành thấp hoặc trung bình. Và những căn hộ không gần bất kỳ trạm dừng nào, đều là những căn hộ trẻ.

5 Kết luận

Qua việc trực quan những features quan trọng của bộ dữ liệu, ta thấy rằng

- Nhìn chung người dân tại phường Daebong, thành phố Daegu, Hàn Quốc có xu hướng chọn mua những căn hộ gần các phương tiện công cộng và ưu tiên chọn những căn hộ có tuổi thọ thấp mặc dù giá thành khá cao.
- Giá giao dịch các căn hộ cũng có xu hướng tăng dần theo thời gian. Cụ thể, số lượng căn hộ giao dịch ngày càng tăng so với trước đó và những căn hộ có kích thước lớn (trên 1000 square feet) cũng được chú ý nhiều hơn. Có thể nói mức sống của người dân tại địa phương này đã được cải thiện khiến cho giá thành tăng cao.
- Số lượng căn hộ có kích thước vừa và nhỏ tại có xu hướng tăng mạnh có thể nói lên được người dân có xu hướng lựa chọn những căn hộ phù hợp với kinh tế của bản thân và có không gian vừa đủ cho gia đình của mình.
- Thị trường mua bán các căn hộ mặc dù có nhiều biến động trong giá thành và số lượng giao dịch. Tuy nhiên, nhìn chung vẫn có xu hướng tăng so với những năm trước đó