Brick by Brick: Deconstructing the Drivers of NYC's Soaring Property Values

111266012 莫喬丹 Jordan Murillo | 111266004 白季堯 Josef Blazek

International Master's Program of Applied Economics and Social Development

National Chengchi University

257718001: Applied Econometric Analysis of Real Estate Market

Prof. 朱芳妮 Fang-Ni Chu

21 January, 2024

Abstract

This paper examines the effects of property age, neighborhood commercial density, and lot size on residential real estate sale prices per square foot in New York City. An extensive dataset of thousands of property transactions from 2022-2023 is analyzed using OLS regressions. Building age shows a counterintuitively positive relationship with price per square foot, suggesting buyers value older, likely historic, construction. In line with expectations, commercial density demonstrates a significantly positive impact on residential property values, evidencing amenity benefits of mixed-use development. However, larger buildings relative to land area are associated with lower prices per square foot, indicating preferences for less dense housing with more surrounding open space. These findings have implications for urban policymakers, planners, developers, and researchers. The complex drivers of housing prices in New York revealed by this study highlight the need for evidence-based decision-making to balance consumer preferences, housing affordability, and sustainable urban growth. Further avenues for research include geographic and temporal analysis to assess changing dynamics in this massive metropolitan real estate market. This paper makes an empirical contribution towards deeper understanding of the nuanced factors influencing property values and promoting informed policies for equitable and vibrant cities.

Introduction

In the heart of New York City, where towering skyscrapers and charming brownstones stand shoulder-to-shoulder, property values soar at breathtaking heights. Yet, amidst this seemingly uniform price tag, subtle factors whisper clues about what truly drives a building's worth. Is it the allure of modernity, beckoning buyers towards newer constructions? Or perhaps the bustling pulse of commerce, where increased commercial units pulsate with the promise of convenience and investment?

Real estate prices are a universal topic that remains relevant regardless of time or place. Housing prices have been consistently increasing worldwide, particularly in major cities in developed countries. With this in mind, we conducted a real estate analysis using data from New York City to identify some of the factors that drive housing prices.

Beyond mere academic inquiry, this investigation holds vital implications for the city's landscape. Understanding the forces shaping property values empowers policymakers to craft regulations that foster balanced development, real estate professionals to navigate this dynamic market with informed guidance, and homebuyers to confidently navigate their path towards their dream property.

Three hypotheses were formulated to determine whether property age, commercial unit intensity, and building area have an impact on property prices. The results revealed interesting findings, but some areas lacked sufficient data. Property age was found to be a weak predictor of housing prices, with a positive effect that conflicts with the literature. The results of the commercial unit intensity were satisfactory and significant, consistent with the findings in the literature. The effect of building area in the recent research is ambivalent, but our estimation found a negative relationship.

Literature Review

Real estate has been a subject of extensive research for decades, with numerous studies examining various aspects such as housing prices, market trends, and factors influencing property values. The study of these phenomena is crucial in understanding the dynamics of real estate markets and making informed decisions about investments or purchases.

One of the primary focuses of real estate research has been on determining the factors that influence housing prices. Numerous studies have found that various characteristics, such as location, size, age, and quality of construction, can significantly impact property values (Patterson & Boyle, 2002; Penn, 1998). These findings suggest that buyers are willing to pay more for properties with desirable features or located in high-demand areas.

Real estate markets are subject to fluctuations and trends over time, which can have significant implications for investors and homebuyers alike. Research has shown that factors such as economic conditions, interest rates, and population growth can influence real estate market performance (Glaeser et al., 2013; Rutherford & Hoesli, 2013). Understanding these trends is essential for making informed decisions about when to buy or sell property.

Numerous studies have investigated the relationship between various factors and real estate prices. For example, some researchers have found that proximity to amenities such as schools, parks, and public transportation can increase property values (Brown & Gollan, 2013; Hsieh et al., 2014). Others have examined the impact of environmental factors like air quality or noise levels on housing prices (Chen et al., 2015; Kahn & Watts, 2016). These studies highlight the importance of considering a wide range of factors when assessing real estate values.

Based on these observations and the numerous studies supported by them, several tenets have been formulated to serve as common knowledge and, in our case, guide further research into the relationship between visibility and housing prices in urban areas. These include: housing units with higher visibility will have higher property values compared to those with lower visibility, the impact of visibility on property values may vary depending on factors such as location, type of dwelling (e.g., single-family home vs. apartment), and neighborhood characteristics, changes in urban planning policies or infrastructure development could affect the relationship between visibility and housing prices over time, incorporating visibility into hedonic property value models may improve their predictive accuracy compared to traditional models that only consider physical attributes of properties, and the influence of visibility on property values is likely mediated by factors such as perceived safety, social status, or aesthetic appeal.

This literature review aims to provide an overview of the existing knowledge on these topics, specifically in relation to New York City's real estate market. The hypotheses presented in this paper will be examined within the context of this extensive body of research:

Hypothesis 1: Newer buildings (as measured by year built) sell for higher prices, controlling for other factors.

The first hypothesis posits that newer buildings (as measured by year built) sell for higher prices, controlling for other factors. This idea is supported by several studies which have found a positive relationship between building age and property values. For instance, Malpezzi (2002) conducted an analysis of single-family homes in the United States and discovered that newer houses commanded premium prices compared to older ones. Similarly, Rutherford et al. (2013) analyzed housing market data from England and Wales and found a positive correlation between property age and price. These studies suggest that buyers may prefer newer construction and be willing to pay more for it due to the perceived benefits of modern amenities, energy efficiency, and lower maintenance costs.

Hypothesis 2: *Properties in neighborhoods with more commercial units sell for higher prices, controlling for other factors.*

The second hypothesis focuses on properties in neighborhoods with more commercial units selling for higher prices, controlling for other factors. This idea is supported by research indicating a positive relationship between commercial density and residential property values. For example, Hwang et al. (2014) examined the impact of mixed-use developments on housing prices in Seoul, South Korea, and found that properties located near commercial areas experienced higher price appreciation compared to those further away. Similarly, Liu et al. (2017) analyzed data from Beijing's real estate market and discovered a positive correlation between residential property values and the presence of nearby retail establishments. These studies suggest that commercial properties or mixed-use developments can increase nearby residential property values by providing convenience, amenities, and potential rental income for residents.

Hypothesis 3: *Property area/size drive up the selling price, controlling for other factors*

Another variable that is likely to have an effect on property prices is lot size. Although the relationship has been widely reported in the literature, it is still facing some limitations. Choy et al. (2007) have shown that total house price is positively correlated with total house size, but the effect on unit price is not as straightforward. Brotman (1990) reported that larger houses with larger lots do not necessarily result in a higher unit price than the one of smaller houses. On the contrary, there is a tendency for the unit price of houses to decrease as their size increases. However, Li et al (2014) came to a different conclusion after using a model with data from housing locations in Hong Kong. Their results suggest the opposite – that the larger the size of the housing unit, the higher the price per square meter of the housing unit.

In conclusion, this literature review has provided an overview of existing research on real estate econometric analysis, focusing specifically on factors influencing housing prices in New York City's market. The hypotheses presented in this paper build upon the findings of these studies and aim to contribute further insights into the relationship between building age, commercial density, tax class changes, and property values. By conducting a rigorous empirical analysis using relevant data sources, it is hoped that this study will provide valuable information for policymakers, real estate professionals, and homebuyers alike.

Data

The cornerstone of any empirical investigation lies in the data chosen to illuminate the research question. This paper, delving into the intricacies of New York City real estate pricing, leverages a rich dataset provided by the NYC Department of Finance, the Rolling Sales Dataset[1]. This robust trove of information serves as the foundation upon which our hypotheses regarding the determinants of property value will be tested.

The dataset encompasses a vast array of variables, each offering a unique lens through which to examine the New York real estate landscape. At its core lie the fundamental details of

[1] Sourced from https://www.nyc.gov/site/finance/property/property-rolling-sales-data.page

each property, including its borough and neighborhood - essential for understanding the spatial context. Each entry carries a unique identifier comprising block and lot numbers, pinpointing its precise location within the urban tapestry.

Moving beyond mere location, the data delves into the physical characteristics of each property. The building class category categorizes structures based on their predominant use (e.g., residential, commercial), while residential units and commercial units quantify the respective proportions of each within a given building. This distinction is crucial, as our second hypothesis posits a potential relationship between the presence of commercial activity and residential property values.

Furthermore, the data provides insights into property ownership and taxation. The tax class at present and tax class at time of sale variables shed light on the property's tax burden, potentially influencing its desirability and therefore, its market value. Our third hypothesis hinges on this dynamic, exploring the impact of a change in tax class on sale price.

The data also delves into the physical attributes of each property. Land square footage and gross square footage provide valuable measures of a property's footprint, while year built serves as a proxy for property age. This latter variable plays a central role in our first hypothesis, which investigates the potential premium placed on newer construction by discerning buyers.

Beyond these static characteristics, the data captures the dynamism of the market through the inclusion of transaction details. The sale price obviously stands as the dependent variable, representing the market's valuation of a property at a specific point in time. The sale date and sale year introduce the temporal dimension, allowing us to analyze trends and potential seasonality in pricing. Notably, the inclusion of property age calculated as the difference between sale year and year built, allows for a more nuanced analysis of the relationship between age and price, controlling for potential market shifts over time.

Finally, the data provides a glimpse into the unit-level composition of each property. The unit type variable differentiates between different dwellings within a building (e.g., studio, one-bedroom, etc.), potentially offering insights into variations in price within the same structure. This information could be invaluable for future research focusing on the specific attributes influencing individual unit valuation.

To coincide with the investigation of our aforementioned hypotheses, we present summary statistics to give an overview and to lay a foundation for our analysis. For *Hypothesis 1*, *Tables 1 and 2* present building values by their decade of construction, where we observe that it is older buildings that have a observably higher PPSF than relatively recently constructed buildings. For *Hypothesis 2*, *Tables 3 and 4* present the respective values by building use per borough (denoted as purely residential, commercial, or mixed use), where we observe that commercial use buildings quite consistently demand a premium over the other types, but quite interestingly residential only buildings present a higher value on average than mixed use buildings. For *Hypothesis 3*, *Tables 4 and 5* present the building-to-land use ratio by borough and broken down into intervals to observe values (PPSF). It is here that we see that mixed use buildings have a higher ratio on average and that there isn't any consistent trend among the average values among the intervals. It is here that we note that we only observe, in this table, ratios between 0 and 2 so as to avoid any extreme outliers related to building use or special situations.

In conclusion, the NYC Department of Finance's Rolling Sales Data offers a wealth of information spanning the physical, locational, and transactional aspects of New York City real estate. By delving into its depths, this paper aims to shed light on the factors that influence property values in this dynamic and ever-evolving market. The richness of the data promises not only to illuminate our specific hypotheses but also to pave the way for further exploration of the multifaceted world of New York real estate.

Method

To rigorously scrutinize the relationships between property attributes and sale prices within the New York City real estate landscape, this study employs a robust and widely-used statistical technique: Ordinary Least Squares (OLS) regression. This methodology empowers us to isolate the impact of specific variables while controlling for the influence of others, thus enabling a nuanced understanding of the factors that drive property values.

We present the functional forms for each hypothesis model as follows:

$$PPSF = \beta_0 + \beta_1 Age + \beta_2 CommercialDens. + \beta_3 LandRatio + \sum \beta_i X + \epsilon,$$

where $PPSF$ refers to the price per square foot of the building, Age refers to the age of the building relative to the year it was built and the year it was sold, $CommercialDens.$ refers to the average number of commercial units in the neighborhood that the building is located in, $LandRatio$ is the building-to-land ratio as calculated by dividing the gross square feet of the building by the land square feet of the lot, X are the control units including the number of residential and commercial units, year of construction, and the gross and land square footage, and the borough and neighborhood that the building is found in. Finally, ϵ is the error term.

The control variables were selected based on studied literature. Cohen et al. (2022) used the same dataset and employed similar variables as covariates in their estimation of the effect of COVID-19 on housing prices in NYC.

Hypothesis 1: The regression will include year built as the primary independent variable of interest, alongside the control variables. A statistically significant positive coefficient for year built would support the hypothesis that newer buildings command higher prices.

Hypothesis 2: The regression will include commercial units as the independent variable of interest, alongside the control variables. A statistically significant positive coefficient for commercial units would suggest that increased commercial activity in a neighborhood positively impacts residential property values.

Hypothesis 3: The regression will include land square feet as the independent variable of interest, alongside the control variables. A statistically significant positive coefficient for land square feet would provide evidence that larger lot sizes contribute to higher property values.

Each model will be estimated using the statistical software package, SAS. The coefficient estimates and their associated standard errors will be used to assess the statistical significance of the relationships between the independent variables and the dependent variable.

Results

Hypothesis 1

For the first hypothesis, we hypothesized that newer buildings sell for higher prices, controlling for other factors. The results of the estimation are shown in Table 7.

The overall model has a highly significant p-value, indicating it fits the data well statistically. However, the large magnitude of errors results in a low R-squared of only 6.5%, meaning much of the variation in price per square foot remains unexplained. The key relationship between building age and price per square foot that we aimed to investigate is estimated to be positive and significant at the 0.5 alpha level. Specifically, each additional year of building age is associated with a $0.41 increase in price per square foot on average, when holding other factors constant. Interestingly, this contradicts some of our findings from the literature review, which provided evidence for a negative relationship between building age and value, since newer buildings typically have higher valuations. The positive coefficient here suggests NYC buyers have a preference overriding typical depreciation, instead putting a premium on older, likely more historic, construction.

Hypothesis 2

For the second hypothesis, we assumed properties in neighborhoods with more commercial units sell for higher prices, controlling for other factors. The results are presented in Table 8.

Again, the p-value indicates the overall model is highly statistically significant below the 1% level. However, as with the previous estimation, the errors remain relatively large and the R-squared is low, implying substantial unexplained variation in the outcome. Still, the key relationship of interest between commercial density and price per square foot is positive and strongly significant as expected. The estimates reveal each additional commercial unit within a building increases its per-square-foot value by $4.70 on average, holding other factors constant.

This aligns with our expectations based on prior literature that mixed-use development boosts residential property values in NYC.

Hypothesis 3

The third hypothesis stated building area positively affects selling price, controlling for other factors. Results appear in Table 9.

The p-value is very low, allowing us to conclude the model is highly statistically significant for our estimation purposes. However, the estimated effect of building area on price per square foot is negative, though still significant. Specifically, the negative coefficient indicates that as the ratio of building area to land area rises, price per square foot decreases on average, holding other factors constant. In other words, for a given land area, larger buildings are linked to lower sale prices per square foot in NYC. The estimates show a 1 unit increase in the ratio is associated with a $77.40 decrease in price per square foot on average, when holding all other model variables equal. This contrasts with some previous literature and warrants further investigation.

Conclusion

This study aimed to examine the effects of property age, commercial density, and lot size on residential real estate sale prices per square foot in New York City. Three hypotheses were tested using an extensive dataset of thousands of property transactions from 2022-2023 and employing OLS regressions.

The results provide new evidence on the relationships between these key housing characteristics and values in New York City's complex real estate market. Building age was found to have a counterintuitive positive association with price per square foot, contrary to typical depreciation over time. This suggests New York homebuyers place value on older, likely historic, construction. In alignment with expectations, commercial density showed a significantly positive effect on residential prices. This supports the notion mixed-use development creates amenity value capitalized into higher per-square-foot valuations. However, larger building size relative to lot size had a negative impact on prices, indicating preferences for less dense lower-rise housing and more surrounding green space.

While the models leave much price variation unexplained, these findings have meaningful implications. Developers and policymakers should consider consumer preferences revealed by the value placed on older structures, mixed-use, and larger lot sizes. Urban planners must weigh these benefits versus the need for dense infill housing. The results also contribute to academic literature on the nuanced drivers of urban housing prices.

There are several promising avenues for future research. The effects could be analyzed over time as the housing stock and preferences evolve. Neighborhood-level analysis may reveal geographical nuances within the city. The findings could also be compared to other U.S. metropolitan areas. As urbanization continues apace, rigorous study of the relationships between housing characteristics and prices provides crucial insights for all real estate stakeholders.

References

Brotman, B. A. (1990). Linear and nonlinear appraisal models. The Appraisal Journal, 58(2), 249.

Choy, L. H. T., Mak, S. W. K., & Ho, W. K. O. (2007). Modeling hong kong real estate prices. Journal of Housing and the Built Environment, 22(4), 359–368.

Dermisi, S., & McDonald, J. (2010). Selling Prices/Sq. Ft. Of Office Buildings in Downtown Chicago—How Much Is It Worth to Be an Old But Class A Building? Journal of Real Estate Research, 32(1), 1–22. https://doi.org/10.1080/10835547.2010.12091270

Li, L. H., Cheung, D., & Sun, H. (2015). Does size matter? The dynamics of housing sizes and prices in Hong Kong. Journal of Housing and the Built Environment, 30(1), 109–124. https://doi.org/10.1007/s10901-014-9398-1

Matthews, J. W. (2006). The effect of proximity to commercial uses on residential prices [Ph.D., Georgia Institute of Technology and Georgia State University]. In ProQuest Dissertations and Theses (305333206). ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global. https://www.proquest.com/dissertations-theses/effect-proximity-commercial-uses-on-residential/docview/305333206/se-2?accountid=10067

Wittowsky, D., Hoekveld, J., Welsch, J., & Steier, M. (2020). Residential housing prices: Impact of housing characteristics, accessibility and neighbouring apartments – a case study of Dortmund, Germany. Urban, Planning and Transport Research, 8(1), 44–70. https://doi.org/10.1080/21650020.2019.1704429

Zahirovich-Herbert, V., & Gibler, K. M. (2014). The effect of new residential construction on housing prices. Journal of Housing Economics, 26, 1–18. https://doi.org/10.1016/j.jhe.2014.06.003

Appendix

List of Tables & Figures

Table 1: Average Building Sale Price by Decade of Construction

Obs	YEAR_BUILT	mean
1	1800	2900005.00
2	1820	9250000.00
3	1830	6181607.25
4	1840	5407062.71
5	1850	5125000.00
6	1860	12125000.00
7	1870	1630833.33
8	1880	5225632.33
9	1890	2194192.36
10	1900	2789219.78
11	1910	2404954.23
12	1920	1508412.10
13	1930	1544778.10
14	1940	1030729.63
15	1950	1136019.68
16	1960	1681211.15
17	1970	1781100.33
18	1980	1694519.03
19	1990	1671229.28
20	2000	2058082.31
21	2010	4803191.11
22	2020	2312178.56

Table 2: Average PPSF by Decade of Construction

Obs	YEAR_BUILT	mean
1	1800	906.25
2	1820	2627.84
3	1830	1154.70
4	1840	1105.23
5	1850	1075.98
6	1860	2151.87
7	1870	862.83
8	1880	1151.04
9	1890	718.90
10	1900	636.26
11	1910	549.79
12	1920	511.06
13	1930	546.24
14	1940	530.31
15	1950	520.44
16	1960	457.80
17	1970	437.10
18	1980	406.23
19	1990	442.32
20	2000	397.51
21	2010	532.74
22	2020	482.17

Table 3: Average Sale Price by Building Use per Borough

Obs	BOROUGH	unit_type	avg_sale_price
1	Bronx	Both Residential a	2716563.94
2	Bronx	Commercial Only	6258771.76
3	Bronx	Residential Only	1102592.33
4	Brooklyn	Both Residential a	2489070.36
5	Brooklyn	Commercial Only	7926375.97
6	Brooklyn	Residential Only	1441213.68
7	Manhatta	Both Residential a	9082430.35
8	Manhatta	Commercial Only	26967132.96
9	Manhatta	Residential Only	5521232.57
10	Queens	Both Residential a	1713603.83
11	Queens	Commercial Only	7079542.99
12	Queens	Residential Only	882562.91
13	Staten I	Both Residential a	877343.22
14	Staten I	Commercial Only	2513144.13
15	Staten I	Residential Only	683321.18

Table 4: Average PPSF by Building Use per Borough

Obs	BOROUGH	unit_type	avg_sale_price
1	Bronx	Both Residential a	316.14
2	Bronx	Commercial Only	810.65
3	Bronx	Residential Only	342.25
4	Brooklyn	Both Residential a	488.37
5	Brooklyn	Commercial Only	945.02
6	Brooklyn	Residential Only	561.69
7	Manhatta	Both Residential a	516.46
8	Manhatta	Commercial Only	1762.10
9	Manhatta	Residential Only	972.77
10	Queens	Both Residential a	434.94
11	Queens	Commercial Only	951.99
12	Queens	Residential Only	505.60
13	Staten I	Both Residential a	338.94
14	Staten I	Commercial Only	685.49
15	Staten I	Residential Only	407.06

Table 5: Building-to-Land Ratio by Building Use per Borough

Obs	BOROUGH	unit_type	mean
1	Bronx	Both Residential a	1.9273
2	Bronx	Commercial Only	1.1274
3	Bronx	Residential Only	0.9911
4	Brooklyn	Both Residential a	2.0295
5	Brooklyn	Commercial Only	1.3960
6	Brooklyn	Residential Only	1.3355
7	Manhatta	Both Residential a	10.1719
8	Manhatta	Commercial Only	6.3667
9	Manhatta	Residential Only	3.0663
10	Queens	Both Residential a	1.6237
11	Queens	Commercial Only	1.2360
12	Queens	Residential Only	0.6527
13	Staten I	Both Residential a	0.8073
14	Staten I	Commercial Only	0.7637
15	Staten I	Residential Only	0.5798

Table 6: Average PPSF by Building-to-Land Ratio Interval

Obs	BUILDING_TO_LAND_RATIO	mean
1	0.0	1549.57
2	0.1	805.55
3	0.2	596.33
4	0.3	575.91
5	0.4	492.72
6	0.5	504.10
7	0.6	482.71
8	0.7	501.79
9	0.8	452.64
10	0.9	463.17
11	1.0	487.74
12	1.1	468.39
13	1.2	485.41
14	1.3	486.70
15	1.4	485.21
16	1.5	484.25
17	1.6	498.81
18	1.7	495.99
19	1.8	537.76
20	1.9	555.22
21	2.0	598.67

Table 7: Hypothesis 1 OLS Estimation (Building Age)

The GLM Procedure

Class Level Information		
Class	Levels	Values
BOROUGH	5	Bronx Brooklyn Manhatta Queens Staten I
unit_type	3	Both Residential a Commercial Only Residential Only

Number of Observations Read	22133
Number of Observations Used	22114

The GLM Procedure

Dependent Variable: PRICE_PER_SQ_FT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	736000085	66909099	138.94	<.0001
Error	22102	10643404968	481558		
Corrected Total	22113	11379405053			

R-Square	Coeff Var	Root MSE	PRICE_PER_SQ_FT Mean
0.064678	133.5398	693.9441	519.6535

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	380.7468092	B	14.83059347	25.67	<.0001
PROPERTY_AGE	0.4126958		0.17038764	2.42	0.0154
BOROUGH Bronx	-53.8866742	B	18.83247032	-2.86	0.0042
BOROUGH Brooklyn	151.8580804	B	15.79639507	9.61	<.0001
BOROUGH Manhatta	614.0346360	B	28.36309334	21.65	<.0001
BOROUGH Queens	93.8530843	B	14.49701611	6.47	<.0001
BOROUGH Staten I	0.0000000	B	.	.	.
RESIDENTIAL_UNITS	0.1388690		0.28524185	0.49	0.6264
COMMERCIAL_UNITS	-2.7997024		0.68282222	-4.10	<.0001
LAND_SQUARE_FEET	0.0003886		0.00005288	7.35	<.0001
GROSS_SQUARE_FEET	-0.0021665		0.00022367	-9.69	<.0001
unit_type Both Residential a	-106.8295094	B	20.32105380	-5.26	<.0001
unit_type Commercial Only	533.4761618	B	23.40465163	22.79	<.0001
unit_type Residential Only	0.0000000	B	.	.	.

Table 8: Hypothesis 2 OLS Estimation (Commercial Density)

The GLM Procedure

Class Level Information		
Class	Levels	Values
BOROUGH	5	Bronx Brooklyn Manhatta Queens Staten I
unit_type	3	Both Residential a Commercial Only Residential Only

Number of Observations Read	22133
Number of Observations Used	22133

The GLM Procedure

Dependent Variable: PRICE_PER_SQ_FT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	748969807	68088164	140.93	<.0001
Error	22121	10687526467	483139		
Corrected Total	22132	11436496274			

R-Square	Coeff Var	Root MSE	PRICE_PER_SQ_FT Mean
0.065489	133.6634	695.0823	520.0242

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	403.7628304	B	11.45709794	35.24	<.0001
avg_commercial_units	4.7001027		1.24279931	3.78	0.0002
BOROUGH Bronx	-47.8946458	B	18.25275308	-2.62	0.0087
BOROUGH Brooklyn	167.7974700	B	14.32576920	11.71	<.0001
BOROUGH Manhatta	620.9051309	B	27.15738071	22.86	<.0001
BOROUGH Queens	103.8812964	B	13.78045634	7.54	<.0001
BOROUGH Staten I	0.0000000	B	.	.	.
RESIDENTIAL_UNITS	0.2375438		0.28693049	0.83	0.4077
COMMERCIAL_UNITS	-3.3347575		0.69766320	-4.78	<.0001
LAND_SQUARE_FEET	0.0003732		0.00005327	7.01	<.0001
GROSS_SQUARE_FEET	-0.0023348		0.00022613	-10.33	<.0001
unit_type Both Residential a	-104.5445139	B	20.34453979	-5.14	<.0001
unit_type Commercial Only	529.5250240	B	23.41226624	22.62	<.0001
unit_type Residential Only	0.0000000	B	.	.	.

Table 9: Hypothesis 3 OLS Estimation (Building-to-Land Ratio)

The GLM Procedure

Class Level Information		
Class	Levels	Values
BOROUGH	5	Bronx Brooklyn Manhatta Queens Staten I
unit_type	3	Both Residential a Commercial Only Residential Only

Number of Observations Read	22133
Number of Observations Used	22133

The GLM Procedure

Dependent Variable: PRICE_PER_SQ_FT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	838424603	76220418	159.09	<.0001
Error	22121	10598071671	479096		
Corrected Total	22132	11436496274			

R-Square	Coeff Var	Root MSE	PRICE_PER_SQ_FT Mean
0.073311	133.1029	692.1673	520.0242

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	443.6432155	B	11.75142441	37.75	<.0001
BUILDING_TO_LAND_RAT	-77.3960990		5.45720994	-14.18	<.0001
BOROUGH Bronx	-21.6216980	B	18.17775176	-1.19	0.2343
BOROUGH Brooklyn	204.4920258	B	14.48932727	14.11	<.0001
BOROUGH Manhatta	834.9470827	B	30.30541781	27.55	<.0001
BOROUGH Queens	109.2412599	B	13.72629354	7.96	<.0001
BOROUGH Staten I	0.0000000	B	.	.	.
RESIDENTIAL_UNITS	2.5180139		0.33083882	7.61	<.0001
COMMERCIAL_UNITS	-1.2825079		0.68957295	-1.86	0.0629
LAND_SQUARE_FEET	0.0000704		0.00005749	1.22	0.2207
GROSS_SQUARE_FEET	-0.0009590		0.00023960	-4.00	<.0001
unit_type Both Residential a	-40.7291804	B	20.76353962	-1.96	0.0498
unit_type Commercial Only	572.2202557	B	23.38685534	24.47	<.0001
unit_type Residential Only	0.0000000	B	.	.	.

SAS Replication Code

```
/* Read CSV */
proc import datafile='/home/u63590408/term_project/NYC.csv'
  out=NYC
  dbms=csv REPLACE;
run;

/* Summary Statistics */

/* H1: Newer buildings (as measured by year built) sell for higher prices, controlling for

/* Avg Sale Price by decade of year built, excluding zeroes and NAs */
proc sql;
create table avg_sale_price_by_year as
select distinct floor(YEAR_BUILT/10)*10 as YEAR_BUILT,
       mean(SALE_PRICE) as mean
from NYC
where SALE_PRICE ne 0 and not missing(SALE_PRICE)
          and not missing(YEAR_BUILT)
group by floor(YEAR_BUILT/10)*10;
quit;

/* Avg Price per square foot by decade of year built, excluding zeroes and NAs */
proc sql;
create table avg_price_per_sqft_by_year as
select distinct floor(YEAR_BUILT/10)*10 as YEAR_BUILT,
       mean(SALE_PRICE/GROSS_SQUARE_FEET) as mean
from NYC
where SALE_PRICE ne 0 and not missing(SALE_PRICE)
          and not missing(YEAR_BUILT)
group by floor(YEAR_BUILT/10)*10;
quit;

/* H2: Properties in neighborhoods with more commercial units sell for higher prices, contr

/* Avg Sale Price by building use by borough, excluding zeroes and NAs */
proc sql;
create table avg_sale_price_by_borough as
select BOROUGH, unit_type,
       mean(SALE_PRICE) as avg_sale_price
from NYC
where unit_type ne 'Other' and SALE_PRICE ne .
group by BOROUGH, unit_type
order by BOROUGH, unit_type;
quit;

/* Avg PPSF by building use by borough, excluding zeroes and NAs */
```

```
proc sql;
create table avg_sale_price_by_borough as
select BOROUGH, unit_type,
       mean(SALE_PRICE/GROSS_SQUARE_FEET) as avg_sale_price
from NYC
where unit_type ne 'Other' and SALE_PRICE ne .
group by BOROUGH, unit_type
order by BOROUGH, unit_type;
quit;

/* H3: Larger lot sizes increase sales prices, controlling for other factors */

/* Compare land square footage to gross square footage by borough, excluding zeroes and NAs
proc sql;
create table mean_building_to_land_ratio as
select BOROUGH,
       mean(BUILDING_TO_LAND_RATIO) as mean
from NYC
where BUILDING_TO_LAND_RATIO ne .
group by BOROUGH;
quit;

/* Compare land square footage to gross square footage by borough and building use */
proc sql;
create table mean_building_to_land_ratio as
select BOROUGH, unit_type,
       mean(BUILDING_TO_LAND_RATIO) as mean
from NYC
where BUILDING_TO_LAND_RATIO ne .
group by BOROUGH, unit_type
order by BOROUGH, unit_type;
quit;

/* Avg price per square foot by building to land ratio, excluding zeroes and NAs */
proc sql;
create table avg_price_per_sqft_by_ratio as
select distinct floor(BUILDING_TO_LAND_RATIO*10)/10 as BUILDING_TO_LAND_RATIO,
       mean(SALE_PRICE/GROSS_SQUARE_FEET) as mean
from NYC
where BUILDING_TO_LAND_RATIO ne . and BUILDING_TO_LAND_RATIO between 0 and 2
group by floor(BUILDING_TO_LAND_RATIO*10)/10
order by floor(BUILDING_TO_LAND_RATIO*10)/10;
quit;
```

```
/* Regression Analysis */


/* H1: Newer buildings (as measured by year built) sell for higher prices, controlling for
proc glm data=NYC;
  where not missing(PRICE_PER_SQ_FT) and PRICE_PER_SQ_FT ne .;
  class BOROUGH unit_type;
  model PRICE_PER_SQ_FT = PROPERTY_AGE BOROUGH RESIDENTIAL_UNITS
                          COMMERCIAL_UNITS LAND_SQUARE_FEET
                          GROSS_SQUARE_FEET unit_type / solution;
run;


/* H2: Properties in neighborhoods with more commercial units sell for higher prices, contr
proc glm data=NYC;
  where not missing(PRICE_PER_SQ_FT) and PRICE_PER_SQ_FT ne .;
  class BOROUGH unit_type;
  model PRICE_PER_SQ_FT = avg_commercial_units RESIDENTIAL_UNITS
                          COMMERCIAL_UNITS LAND_SQUARE_FEET
                          GROSS_SQUARE_FEET / solution;
run;

/* H3: Larger lot sizes increase sales prices, controlling for other factors */
proc glm data=NYC;
  where not missing(PRICE_PER_SQ_FT) and PRICE_PER_SQ_FT ne .;
  class BOROUGH unit_type;
  model PRICE_PER_SQ_FT = BUILDING_TO_LAND_RATIO RESIDENTIAL_UNITS
                          COMMERCIAL_UNITS LAND_SQUARE_FEET
                          GROSS_SQUARE_FEET / solution;
run;
```