# Real Estate Appraisal:

# Impact of Marketing Time and Real Estate Characteristics

**CMIS 567-73C**
**Data Analytics Project for Business**

**Submitted to**

Professor Jospeh Vithayathil

**Submitted by:**

Sankita Sigdel

March 6th, 2025

# Table of Contents

# Executive Summary

Corporate Settlement Solutions (CSS) estimated 51% of the real estate sales in the United States had an appraisal value more than the market price. This discrepancy raises a question: Is it the market condition or housing characteristics that are causing appraisal value to be above contract price? Freddie Mac, a Government-Sponsored Enterprise (GSE) has hired me as an independent contractor to answer these questions and examine the underlying factors contributing to this trend using the Unified Appraisal Dataset (UAD). GSEs and real estate investors rely on accurate property valuations to assess risk, inform lending decisions, and guide investment strategies. This analysis will help GSEs to refine their risk assessment models for loan-to-value calculations and update appraisal guidelines to ensure more accurate and consistent property valuations.

The dataset, obtained from the Federal Housing Finance Agency (FHFA), contains information about housing characteristics, economic trends, and appraisers' evaluation from 2015 to 2022. This study utilizes Multiple Linear regression and Logistic Regression, with fixed effects for time and location to explain the influence of real estate characteristics on Marketing Time, and Contract price to Appraisal value Ratio.

The analysis examined key variables including real estate attribute's location, condition, quality and owner occupancy. It revealed that a property - remodeled within 15 years, owner occupied and located in a metro area had a higher likelihood of being sold at a higher price than the appraised value. It indicates that buyers prefer modern, well-maintained properties in high demand urban locations. Moreover, a remodeled property is 17.5% more likely to sell within 3 months than a property not remodeled which suggests real estate investors to remodel properties to sell at a higher selling price and for quicker sales.

The result challenges conventional belief that quality rating has a significant impact on premium pricing of a property. It is unclear how subjective judgement in quality rating impacts the appraisal outcome. Therefore, I recommend GSEs to conduct research on this area and educate appraisers about potential biases in quality assessment and their impact on property valuation. Additionally, I recommend GSEs to regularly conduct audits on appraisal methods to reflect evolving customer preferences such as shifting priorities from quality to aesthetics and remodeled properties.

# Introduction

Real estate valuation is the process of determining property's worth, also known as appraisal value. Property appraisal is done by licensed professionals usually for real estate transactions such as buying, selling, refinancing, and lending. Appraisal value is determined by comparing recent sales, and current market price with a group of properties with similar characteristics. Appraisal is influenced by housing characteristics, location, and condition of the housing market at the time of appraisal. During the seller's market (more buyers than sellers), appraisal of a property will be higher for the same property compared to buyer's market (more sellers than buyers). Contract Price, on another hand, is the value of a property that a buyer has agreed to pay to a seller. Contract Price to Appraisal Value (CTA) ratio determines how close the agreed price is to the professional assessment of market value. The CTA ratio measures if a property is sold at premium or discount to its appraised value.

Over the years, numerous studies have investigated factors which influence customers to pay a premium for a property. A study suggests that quality, location, and sustainable housing characteristics influence premium pricing of the house (Leonard, 2013). A study about premium housing sales in Bexar County from 2009 to 2019 showed an owner occupied property is more likely to be sold at premium than a non-owner-occupied property (Cadena & Thomson, 2020). This finding is backed by another study that examines metropolitan areas to have positive influence on premium pricing (Hagen, 2005). Zillow data shows remodeled property are more likely to be sold at premium than a property which does not include any remodeled information. It shows the potential of remodeling to influence property sales. Additionally, Marketing time is another crucial factor which affects housing prices. A study found dynamic pricing to help property be sold in shorter marketing time (English, 2024).

Additionally, appraisal value is determined from historical sales of comparable properties and involves subjective assessment of real estate characteristics like quality and condition. It may however not fully capture factors that drive buyer preference. This potential gap highlights a need for appraisal methods to elucidate market dynamics and evolving consumer trends in property valuation. Although demand-supply and growth-rate dynamics are disclosed by appraisers in the form, they may still have biases from the past (despite a booming/declining market) (Eriksen, Fout, Palim, & Rosenblat, 2019).

Existing studies on buyers' willingness to pay a premium for a real estate property show that some real estate characteristics may need adjustment to align appraisal value more closely to contract price. Property sold at a premium is a missed opportunity for mortgage lenders as they rely on appraisal value to calculate Loan to Value ratio (LTV). If properties sold at premium are appraised closer to their contract value, mortgage lenders can lend higher loans with more confidence and reduced risk, potentially leading to increased loan approvals and more profit for lenders. Therefore, this research would investigate two main areas - Contract Price to Appraisal Value (CTA) and Marketing Time with an objective to understand how different real estate attributes impact pricing and marketability to help Government Sponsored Enterprises (GSEs) to update guidelines for appraisers. Additionally, this study will help real estate professionals on how to position properties for quicker sales.

## Business Questions

This section explains broad business questions for this study.

A. Which real estate characteristic drives premium pricing of a property?

The CTA ratio measures if a property is sold at premium/discount to its appraised value. Understanding real estate characteristic influence on CTA will help appraisers improve valuation closer to contract value and assist mortgage lenders to assess loan risk more effectively.

B. Does premium pricing have an effect on Marketing Time of the property?

Marketing time is a key factor in real estate transactions for influencing both seller's pricing and buyers' urgency. Understanding the relationship between premium-pricing and marketing time would help real estate investors to position themselves for higher profitability without increasing inventory.

# Data

This section explains sample data, list of variables and hypothesis of research question.

## Sample Data

For this analysis, I have used data from the Uniform Appraisal Dataset (UAD) Appraisal-Level Public Use File (PUF) from the year 2015 to 2022 by the Federal Housing Finance Agency (FHFA). The data is collected by appraisers on Uniform Residential Appraisal Report (URAR) for single family homes and represents 5% of nationally represented deidentified data. FHFA performs stratified simple random sampling to draw samples for each year. I individually combined data from each year to 1,238,268 records. This data includes "refinancing" and "home purchase" data for single-family homes. The scope of this report is to study single-family "home purchase" and understand factors affecting Contract Price to Appraisal Value ratio. Therefore, records are filtered for "Home Purchase" - 627,402 records. 14 records did not have information on contract prices and hence were removed to final 627,388 records for analysis. I calculated contract price to appraisal value percentage using the following formula:

$$\text{CTA} = \frac{Contract\ Price}{Appraised\ Value} * 100$$

Table 1 and Table 2 provide the list of numerical and categorical variables used for this analysis.

| Variables | Acronym | Details | Unit of Measure |
|---|---|---|---|
| Contract Price | CP | Reported contract price of the property | USD |
| Appraised Value | AV | Appraiser specified value of the property | USD |
| Contract price to Appraisal value | CTA | Percentage of Contract Price (CP) relative to Appraised Value (AV) | Derived field in Percentage |

*Table 1: Numeric variables from dataset*

| Variables | Acronym | Details | Level |
|---|---|---|---|
| Quality | Q | Quality rating of the property. Q1 is the highest and Q6 is lowest rating. | Q1, Q2, Q3, Q4, Q5/Q6 |
| Condition | Con | Condition rating of the property. C1 is the highest and C6 is lowest rating. | C1, C2, C3, C4, C5/C6 |
| Owner Occupied | OO | Property occupied by Homeowners. | Yes, No |
| Updated within 15 years | Upd | Remodeling in kitchen or bathroom within 15 years | Yes, No |
| Growth Rate | GR | Growth Rate at which neighborhood is being developed | Slow, Stable, Rapid |
| Bedroom | Bed | Total bedroom on the property. | 0-2 bedrooms, 3 bedrooms, 4+bedrooms |
| Bathroom | Bath | Total bathroom on the property. | 1 Bathroom, 2 Bathrooms, 3 Bathrooms, 4+ Bathrooms |
| Lot Size | LS | Lot size category in acres. | Less than 1/8 acre, 1/8 up to ¼ acre, ¼ up to ½ acre, ½ up to 1 acre, 1+ acre |
| Market Time | MT | Estimated time on market. | Under 3 months, Over 3 months |
| Year | Yr | Year of appraisal report. | 2015 to 2022 |
| State Fips | SFips | Federal code to identify States | 1 to 2 digit character |
| County Fips | CFips | Federal code to identify County. | 3 to 5 digit character |
| Tract Fips | TFips | Federal code to identify Tract. | 6 to 11 digit character |
| Dts_rural | DR | Census tract outside of Metropolitan area. | Yes, No |

*Table 2: Categorical variables from dataset*

## Hypothesis

Real estate characteristics play a pivotal role in premium pricing of a property. This study examines key real estate attributes that impact the Contract Price to Appraisal Value (CTA), with an objective to provide insights on pricing dynamics and market behavior. Therefore, following hypotheses were tested to investigate research objective of this report:

A. Which real estate characteristic affects premium pricing of a property?

*Hypothesis 1: Property remodeled within 15 years is more likely to sell at a premium price.*

CTA = f($Upd_{Yes}$, Bath, Bed , Q , Yr, Tips)

*Hypothesis 2: Property in a metropolitan area is more likely to sell at a premium price than in rural area.*

CTA = f($DR_{No}$, LS, Bath , Bed, Q , Yr, Tips)

*Hypothesis 3: Owner occupied property is more likely to sell at a premium price than non-owner occupied property.*

CTA = f($OO_{Yes}$, Bed, Bath, Q, Yr, Tfips)

B. Does premium pricing have an effect on Marketing Time of the property?

*Hypothesis 4: Property with higher CTA (premium pricing) is likely to sell in less than three months.*

$MT_{<3\ months}$ = f( CTA, Upd, DR, GR , Yr, Tfips)

## Data Visualization

To explore trends in data, scattered charts and bar charts were created in R. Figure 1 below shows a scattered chart to compare the relationship between appraisal value and contract price. There is a linear relationship between contract price and appraisal value. However, there are many data points in the scattered chart showing Contract Price greater than Appraisal Price implying properties being sold in premium compared to appraisal value. This finding will be explored in a later section using regression analysis.

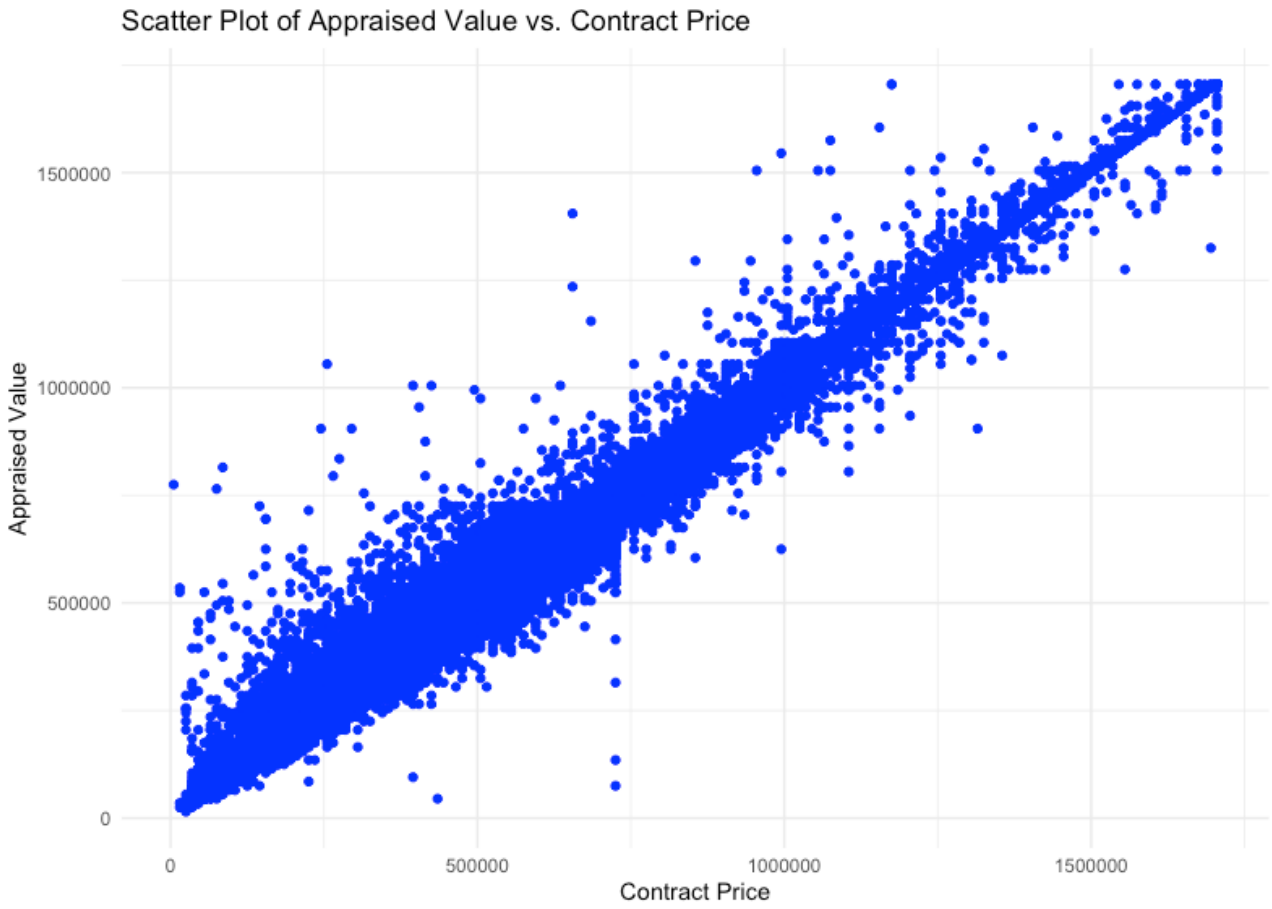


*Figure 1: Scatter plot of Contract Price and Appraisal Value*

Figure 2 investigates the contract price to appraisal value (CTA) ratio to see how many properties were sold at a premium price for years 2015 to 2022. CTA_category "Over 100" shows the number of properties sold at a premium price compared to appraisal value. According to Figure 2, 39,043 properties were sold at premium price between 2015 and 2022. These are missed business opportunities for mortgage lenders. This finding will be explored in a later section using regression analysis.

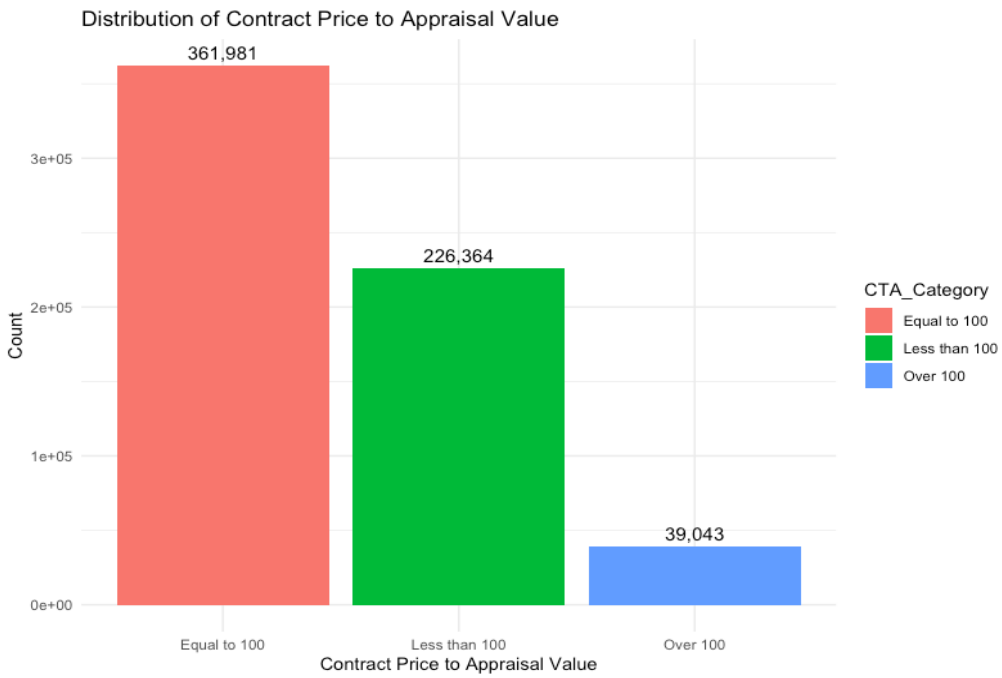


*Figure 2: Distribution of contract price to appraisal value (CTA) ratio*

Figure 3 shows distribution of premium priced property by marketing time. Most properties with premium prices are sold within 3 months. This finding will be explored in a later section using regression analysis.

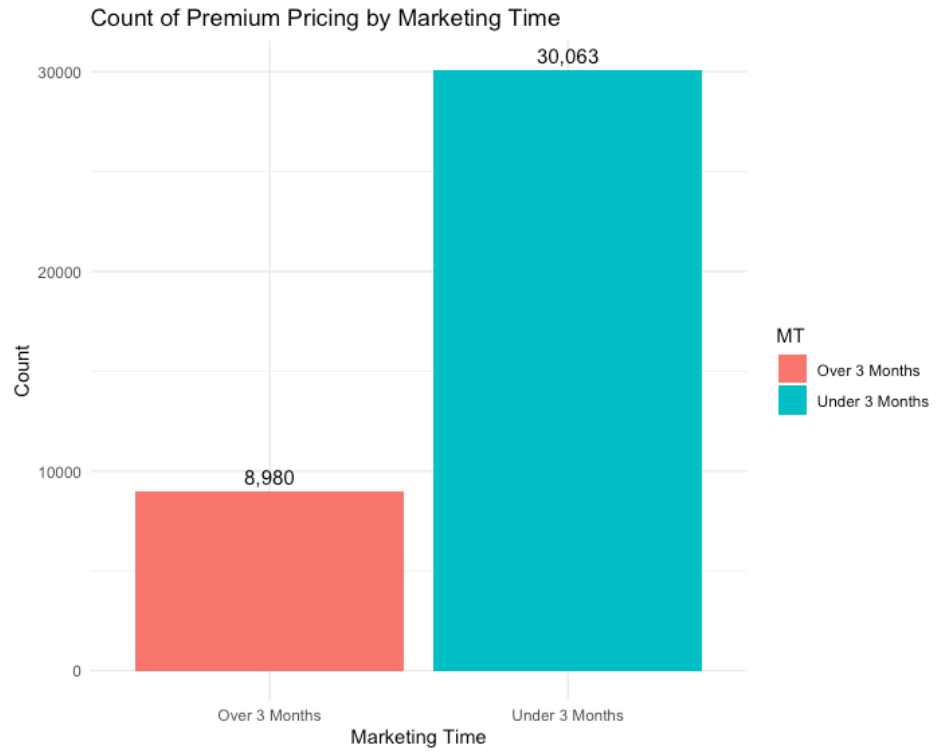


*Figure 3: Distribution of premium priced properties by marketing time*

Figure 4 shows distribution of premium priced property by Owner Occupied. The majority of properties with premium prices are owner occupied. This finding will be explored in a later section using regression analysis.

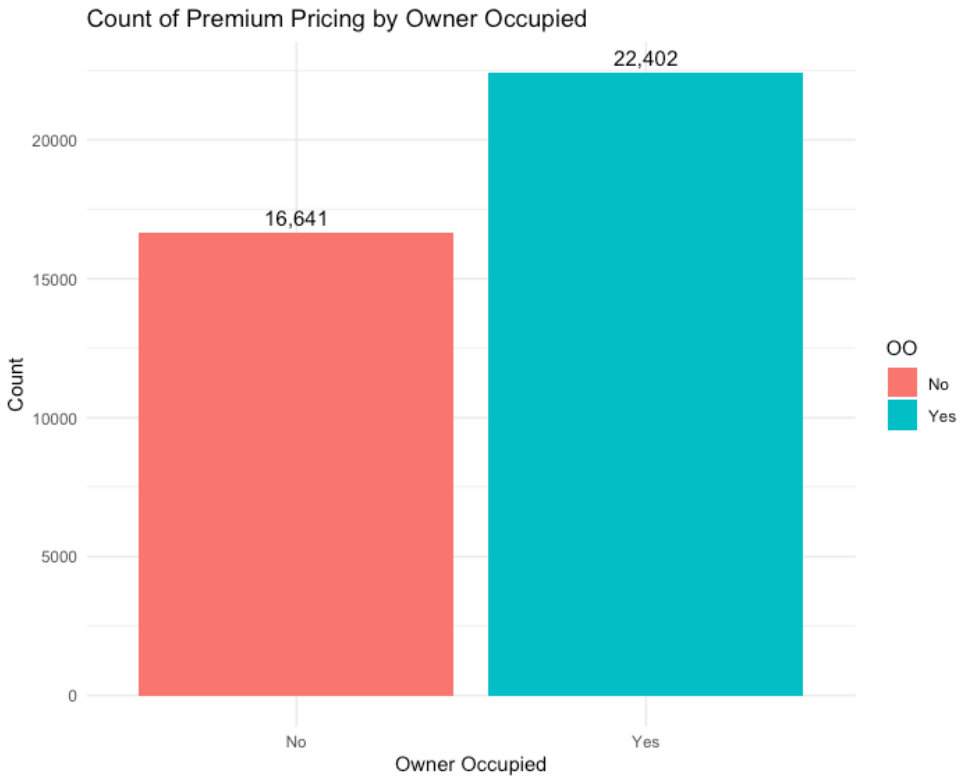


*Figure 4: Distribution of premium priced properties by Owner Occupied*

Figure 5 shows distribution of premium priced property by rural/metro area. Most properties with premium prices are in metropolitan areas. This finding will be explored in a later section using regression analysis.


*Figure 5: Distribution of premium priced properties by Rural Area*

Figure 6 shows distribution of premium priced property by remodeled property. Many properties with premium prices are remodeled properties. This finding will be explored in a later section using regression analysis.

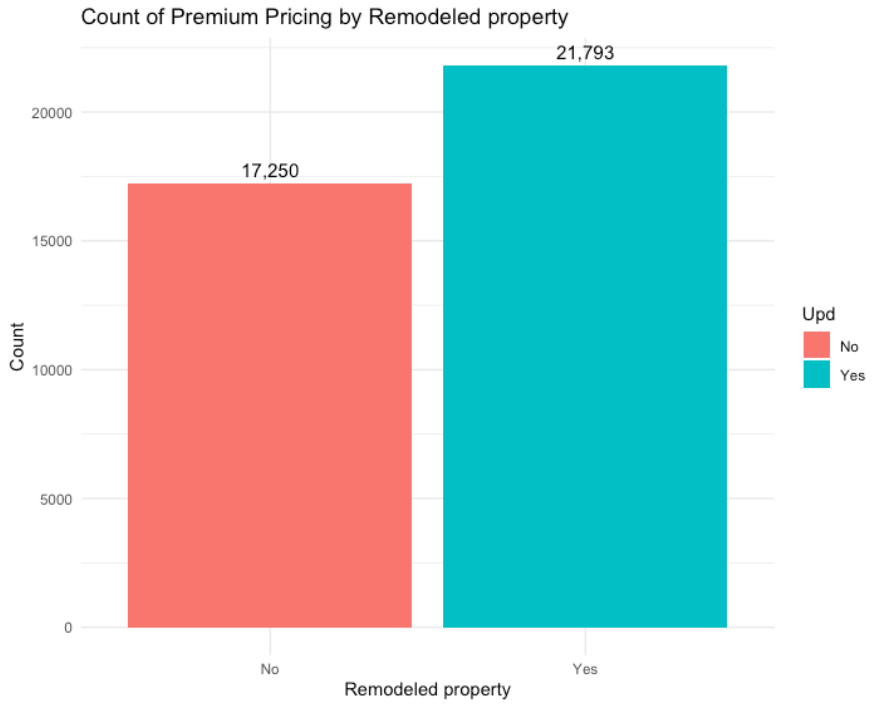


*Figure 6: Distribution of premium priced properties by Remodeled property*

# Methodology

The following regression models are executed to investigate the research objective of this report. One-way within fixed effects is applied to the regression to remove unobserved heterogeneity for county and year-specific factors.

$$CTA_{it} = \beta_{0i} + \beta_1 DR_{No,it} + \sum_{k=1}^{3} \beta_{2k} D_{LS_{k,it}} + \sum_{l=1}^{2} \beta_{3l} D_{Bath_{k,it}} + \sum_{m=1}^{4} \beta_{4m} D_{Bed_{m,it}} + \sum_{m=1}^{4} \beta_{4m} D_{Q_{m,it}} + \alpha_i + \gamma_t + \varepsilon_{it} \quad (1)$$

$$CTA_{it} = \beta_{0i} + \beta_1 Upd_{Yes,it} + \sum_{k=1}^{3} \beta_{2k} D_{LS_{k,it}} + \sum_{l=1}^{2} \beta_{3l} D_{Bath_{k,it}} + \sum_{m=1}^{4} \beta_{4m} D_{Bed_{m,it}} + \sum_{m=1}^{4} \beta_{5m} D_{Q_{m,it}} + \alpha_i + \gamma_t + \varepsilon_{it} \quad (2)$$

$$CTA_{it} = \beta_{0i} + \beta_1 OO_{Yes,it} + + \sum_{l=1}^{2} \beta_{3l} D_{Bath_{k,it}} + \sum_{m=1}^{4} \beta_{4m} D_{Bed_{m,it}} + \sum_{m=1}^{4} \beta_{5m} D_{Q_{m,it}} + \alpha_i + \gamma_t + \varepsilon_{it} \quad (3)$$

$$\text{Logit}(P(MT_{<3months})) = \beta_0 + \beta_1 CTA_{it} + \beta_2 Upd_{yes,it} + \beta_3 DR_{No,it} + \sum_{m=1}^{2} \beta_{5m} D_{GR_{m,it}} + \alpha_i + \gamma_t + \varepsilon_{it} \quad (4)$$

*Where:*

$D_{Q_k}$ = *dummy variables for Q, excluding the reference category.*

$D_{Con_k}$ = *dummy variables for Con, excluding the reference category.*

$D_{GR_k}$ = *dummy variables for GR, excluding the reference category.*

$D_{LS_k}$ = *dummy variables for LS, excluding the reference category.*

$D_{Bed}$ = *dummy variables for Bed, excluding the reference category.*

$D_{Bath}$ = *dummy variables for Bath, excluding the reference category.*

$\alpha_i$ = *county-level fixed effect (county).*

$\gamma_t$ = *time fixed effect (year).*

$\varepsilon_{it}$ = *error term*

# Data Analysis and Findings

This section contains descriptive statistics and regression results.

## Descriptive Statistics

The result presented in Table 3 is a descriptive statistics of numerical variables in the dataset. There is high variability in the minimum and maximum contract price (CP) and appraisal value (AV) of a property. Therefore, it can be inferred that the dataset includes both dilapidated and luxury properties. Additionally, mean and median for Contract Price to Appraisal Value (CTA) is around 100 suggesting property's appraisal value (AV) to be closer to contract price (CP). Meanwhile, CTA minimum value is 0.6 and maximum value is 966.7. Therefore, it can be inferred that the property may be significantly underpriced or overpriced compared to their appraised value.

| Variables | N | Mean | Median | Standard Deviation | Minimum | Maximum |
|---|---|---|---|---|---|---|
| CP | 627,388 | $344,148.09 | $315,000 | $183,429.55 | $5,000 | $1,705,000 |
| AV | 627,388 | $348,727.20 | $315,000 | $184,306.69 | $15,000 | $1,705,000 |
| CTA | 627,388 | 98.52 | 100 | 4.69 | 0.6 | 966.7 |

*Table 3: Descriptive statistics for Numerical Values in Dataset*

## Regression Results

Regression equation is used to explain the real estate characteristics affecting premium pricing of property. Regression models are executed in R.

**A. Which real estate characteristic drives premium pricing of a property?**

This question examines how real estate characteristics like Remodeling within 15 years, Property location and neighborhood growth rate affects Contract to Appraisal Ratio.

*Hypothesis 1: Property remodeled within 15 years is more likely to sell at premium price than not remodeled.*

This hypothesis explains the relationship between remodeling of a property within 15 years (UpdYes) and Contract price to Appraisal value (CTA). To investigate this hypothesis, CTA is a dependent variable, and UpdYes is an independent variable in the regression model. Bed, Bath, and Quality rating are also included as control variables which provide other significant insights. Estimation of the impact of remodeled property on Contract price to Appraisal value is shown in Table 4.

$$CTA_{it} = \beta_{0i} + \beta_1\, Upd_{yes,it} + \sum_{k=1}^{3} \beta_{2k}\, D_{Bath_{k,it}} + \sum_{l=1}^{2} \beta_{3l}\, D_{Bed_{k,it}} + \sum_{m=1}^{4} \beta_{4m}\, D_{Q_{m,it}} + \alpha_i + \gamma_t + \varepsilon_{it}$$

```
Coefficients:
                  Estimate Std. Error t-value  Pr(>|t|)
UpdYes            0.481631   0.011959 40.2730 < 2.2e-16 ***
Bath2 Bathrooms   0.674390   0.022710 29.6962 < 2.2e-16 ***
Bath3 Bathrooms   0.863863   0.024538 35.2057 < 2.2e-16 ***
Bath4+ Bathrooms  0.707365   0.029263 24.1731 < 2.2e-16 ***
Bed3 bedrooms     0.104111   0.020676  5.0353 4.773e-07 ***
Bed4+ bedrooms    0.010514   0.023000  0.4571    0.6476
QQ2               0.073878   0.194573  0.3797    0.7042
QQ3              -0.078638   0.189621 -0.4147    0.6784
QQ4              -0.052208   0.189632 -0.2753    0.7831
QQ5 and Q6       -0.172628   0.199585 -0.8649    0.3871
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

| *N* | *F- Statistics* | *p-value* | *R-squared* | *Adj R-squared* |
|---|---|---|---|---|
| *627,388* | *296.084* | *< 0.001* | *0.005* | *0.005* |

*Table 4: Multiple Linear Regression results for Hypothesis 1*

The R-squared of the regression is 0.005 which means that 0.5% of the variance in dependent variable Contract price to Appraisal value (CTA) is described by independent variables.

Property updated within 15 years (UpdYes) is a significant variable and has a positive coefficient of 0.481, showing positive relationship with dependent variable CTA. It means that a property remodeled within 15 years is likely to increase CTA by 0.481 as compared to a property not remodeled within 15 years while holding other variables constant. This supports a general expectation that a property remodeled within 15 years has a higher CTA than one not remodeled. This may be because people are willing to pay a premium price for a remodeled property which is desirable aesthetically compared to a property not remodeled. Thus, hypothesis 1 is accepted.

Additionally, dummy variable – number of bathrooms is significant with positive coefficient showing it has significant influence on Contract price to Appraisal value (CTA). People are willing to pay higher prices for property with a higher number of bathrooms. However, Quality rating and Bedroom (more than 4) have p-values greater than 0.05. Therefore, they do not have a significant impact on CTA.

*Hypothesis 2: Property in metropolitan areas is more likely to sell at a premium price than in rural areas.*

This hypothesis explains the relationship between the location of a property in a metropolitan area (DRNo) and Contract price to Appraisal value (CTA). To investigate this hypothesis, CTA is the dependent variable, and DRNo is an independent variable in the regression model. Bed, Bath, Lot Size and Quality rating are included as control variables which could provide other significant

insights. Table 5 shows the result for estimating the location impact on Contract price to Appraisal value (CTA).

$$CTA_{it} = \beta_{0i} + \beta_1 DR_{No,it} + \sum_{k=1}^{3} \beta_{2k} D_{LS_{k,it}} + \sum_{l=1}^{2} \beta_{3l} D_{Bath_{k,it}} + \sum_{m=1}^{4} \beta_{4m} D_{Bed_{m,it}} + \sum_{m=1}^{4} \beta_{5m} D_{Q_{m,it}} + \alpha_i + \gamma_t + \varepsilon_{it}$$

```
Coefficients:
                      Estimate Std. Error t-value  Pr(>|t|)
DRNo                  0.725573   0.016786 43.2246 < 2.2e-16 ***
LS1/4 up to 1/2 acre  0.183948   0.024299  7.5701 3.733e-14 ***
LS1/8 up to 1/4 acre  0.291558   0.023065 12.6406 < 2.2e-16 ***
LS1+ acre            -0.236058   0.027650 -8.5374 < 2.2e-16 ***
LSLess than 1/8 acre  0.333587   0.025980 12.8401 < 2.2e-16 ***
Bath2 Bathrooms       0.587027   0.022644 25.9243 < 2.2e-16 ***
Bath3 Bathrooms       0.684857   0.024590 27.8512 < 2.2e-16 ***
Bath4+ Bathrooms      0.513996   0.029338 17.5196 < 2.2e-16 ***
Bed3 bedrooms         0.119039   0.020956  5.6803 1.345e-08 ***
Bed4+ bedrooms        0.040323   0.023897  1.6874   0.09154 .
QQ2                   0.158941   0.194200  0.8184   0.41311
QQ3                  -0.068920   0.189247 -0.3642   0.71572
QQ4                  -0.059877   0.189257 -0.3164   0.75172
QQ5 and Q6           -0.176239   0.199194 -0.8848   0.37628
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

| *N* | *F- Statistics* | *p-value* | *R-squared* | *Adj R-squared* |
|---|---|---|---|---|
| *627,388* | *392.035* | *< 0.001* | *0.008* | *0.008* |

*Table 5: Multiple Linear Regression results for Hypothesis 2*

The R-squared value of the regression is 0.008 which means that 0.8% of the variance in dependent variable Contract price to Appraisal value (CTA) is described by independent variables.

DRNo (Property in metro area) is significant and has a positive coefficient of 0.725 showing a positive relationship with dependent variable CTA. It means that a property in a metropolitan area (DRNo) is likely to increase CTA by 0.725 as compared to a property located in a rural area, holding other variables constant. This supports the expectation that a property in a metropolitan area has a higher CTA than one in a non-rural area. This may be because people are willing to pay a premium price (contract price higher than appraised value) for better opportunities and better amenities in a metropolitan area. Thus, hypothesis 2 is accepted.

Dummy variable - number of bathrooms is significant with positive coefficient showing similar result as hypothesis one. Number of bathrooms is found to have a major influence on Contract price to Appraisal value (CTA). People are willing to pay higher prices for a property with more bathrooms. However, Quality rating and Bedroom (more than 4) have p-values greater than 0.05. Therefore, they do not have a significant impact on CTA.

*Hypothesis 3: Owner occupied property is more likely to sell at a premium price than non-owner occupied property.*

This hypothesis explains the relationship between Owner Occupied property and Contract price to Appraisal value (CTA). To investigate this hypothesis, CTA is a dependent variable, and Owner

Occupied (OO) is an independent variable in the regression model. Bed, Bath and Quality ratings are included as control variables which could provide other significant insights. Table 6 shows the impact of owner occupation (OO) on Contract price to Appraisal value (CTA).

$$CTA_{it} = \beta_{0i} + \beta_1 OO_{No,it} + + \sum_{l=1}^{2} \beta_{3l} D_{Bath_{k,it}} + \sum_{m=1}^{4} \beta_{4m} D_{Bed_{m,it}} + \sum_{m=1}^{4} \beta_{5m} D_{Q_{m,it}} + \alpha_i + \gamma_t + \varepsilon_{it}$$

```
Coefficients:
                  Estimate Std. Error t-value  Pr(>|t|)
OOYes             0.670554   0.011820 56.7328 < 2.2e-16 ***
Bath2 Bathrooms   0.604193   0.022647 26.6784 < 2.2e-16 ***
Bath3 Bathrooms   0.744212   0.024376 30.5299 < 2.2e-16 ***
Bath4+ Bathrooms  0.567649   0.029065 19.5301 < 2.2e-16 ***
Bed3 bedrooms     0.098432   0.020650  4.7666 1.874e-06 ***
Bed4+ bedrooms    0.005458   0.022969  0.2376    0.8122
QQ2              -0.115847   0.194353 -0.5961    0.5511
QQ3              -0.243997   0.189409 -1.2882    0.1977
QQ4              -0.196784   0.189419 -1.0389    0.2989
QQ5 and Q6       -0.301600   0.199348 -1.5129    0.1303
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

| *N* | *F- Statistics* | *p-value* | *R-squared* | *Adj R-squared* |
|---|---|---|---|---|
| *627,388* | *456.094* | *< 0.001* | *0.007* | *0.007* |

*Table 6: Multiple Linear Regression results for Hypothesis 3*

The R-squared of the regression is 0.007 which means that 0.7% of the variance in dependent variable Contract price to Appraisal value (CTA) is described by independent variables.

OOYes (Owner Occupied) is significant and has a positive coefficient of 0.6705 showing positive relationship with dependent variable CTA. It means that the property that is Owner Occupied (OO) is likely to increase CTA by 0.6705 as compared to the property that is not occupied by owners, holding other variables constant. This supports the expectation that a property occupied by the owners is more likely to be sold at a premium price. This may be because owner occupied property has tax benefits. Thus, hypothesis 3 is accepted.

Again, variables - number of bathrooms and bedrooms show a positive relationship. People are willing to pay higher prices for properties with more bathrooms. However, Quality rating and Bedroom categorized as "more than 4" have p-values greater than 0.05. Therefore, we can imply that Quality rating does not impact on CTA.

**B. How does Contract to Appraisal ratio affect Marketing Time of a property?**

This question analyzes the relationship between premium pricing and marketing time.

*Hypothesis 4: Property with higher CTA (premium pricing) is likely to sell in less than three months.*

This hypothesis examines the effect of Contract to Appraisal Ratio (CTA) on the marketing time of a property. To investigate this hypothesis, MT is the binary dependent variable, and CTA is

the independent variable for the regression model. DR, Upd, and GR are included as control variables which could provide other significant insights. Table 7 shows the relationship between Contract price to Appraisal value (CTA) and Marketing Time.

$$\text{Logit}(P(MT_{<3months})) = \beta_0 + \beta_1 CTA_{it} + \beta_2 Upd_{yes,it} + \beta_3 DR_{No,it} + \sum_{m=1}^{2} \beta_{5m} D_{GR_{m,it}} + \alpha_i + \gamma_t + \varepsilon_{it}$$

| | Estimate | Std. Error | z value | Pr(>\|z\|) | | Odd_Ratio |
|---|---|---|---|---|---|---|
| CTA | 0.009016 | 0.001405 | 6.41686 | 1.3911e-10 | *** | 1.0090565 |
| DRNo | 0.426349 | 0.052756 | 8.08158 | 6.3932e-16 | *** | 1.5316546 |
| UpdYes | 0.161143 | 0.011590 | 13.90349 | < 2.2e-16 | *** | 1.1748532 |
| GRSlow | -0.080577 | 0.030059 | -2.68061 | 7.3487e-03 | ** | 0.9225835 |
| GRRapid | 0.762676 | 0.087553 | 8.71103 | < 2.2e-16 | *** | 2.1440060 |

---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

| *N* | *Degree of Freedom* | *Adj. Pseudo R-squared* | *p-value* |
|---|---|---|---|
| *627,388* | *5* | *0.238* | *< 0.001* |

Table 7: Multiple Logistic Regression results for Hypothesis 4

Pseudo R-squared of the regression is 0.238 which means that 23.8% of the variance in MT is described by the model. All independent variables - CTA, DR, Upd and GR have an impact on Marketing Time (MT) of property.

CTA is significant with a p-value less than 0.001 and has a positive coefficient, indicating an increase in CTA will increase the likelihood of Marketing time (MT) less than 3 months. Additionally, the odds ratio of CTA is 1.0091 which means that a unit increase in CTA is associated with 0.91% increase in the odds of a property sold within 3 months. Other independent variables DR, Upd and GR have more impact on marketing time than CTA. Property in metro area (DRNo), remodeled property (UpdYes) and growth rate- "Rapid" (GRRapid) have higher odds of selling within 3 months by 53.2%,17.5% and 114.4% respectively. CTA does not have a significant impact on Marketing time for being sold in less than 3 months. However, a remodeled property in a metro area is more likely to be sold in less than 3 months. Based on the interpretation, hypothesis 4 is accepted.

# Limitation

I have used data from the Federal Housing Finance Agency (FHFA) uniform appraisal dataset for this project. This dataset is a 5% sample drawn randomly from the appraisal population data. Therefore, there could be biases within the samples. Bootstrapping using multiple sets of samples would make the analysis more robust.

Additionally, R-squared values of the regression model are significant but are less than 1%. This shows that other unobserved variables - fluctuation in interest rate, demand and supply imbalances, inflation, mortgage rate and other economic conditions, not included in the models also affect premium pricing. While time and location effects are minimized with fixed effects in the regression, unobserved market conditions are not accounted for.

Independent variables used for this analysis such as marketing time are categorical. Marketing time as a continuous variable would give more confidence as it will capture a wider range of values for precise comparison. Since categorical variables provide distinct groups, there is potential to lose some information within those groups.

## Conclusion and Recommendation

This study examines the real estate characteristics that influence premium pricing, and their impact on marketing time. From the analysis, owner occupancy, metro area location, remodeling, and number of bathrooms emerged as influential factors for premium pricing of a property. Property with more bathrooms is more likely to sell at a premium price than a property with fewer bathrooms. Quality rating of a property does not seem to have a statistical relationship with Contract Price to Appraisal Value ratio, showing it does not drive premium pricing. While premium priced property is more likely to sell quickly, it has limited effect compared to other factors such as property location (metro), remodeled property and rapid neighborhood growth.

The result supported most findings from previous work on factors influencing premium pricing of property (except quality rating). These findings suggest more research is required on factors that people value and are willing to pay extra for in a property. Is quality rating an equally relevant criteria for an appraisal of a property? This should be studied by GSEs and other research teams to adjust appraisal practices.

Based on 2015 to 2022 sample data, quality rating of a property does not seem to influence premium price of a property. This raises a question about how quality is defined and measured in appraisals. Going forward, I would recommend GSEs to conduct a bootstrapping method to analyze multiple samples from the population data to validate regression results from this analysis and adjust real estate characteristics which are not significant for appraisal. By incorporating such checks regularly, they can align the appraisal price closer to the contract price. As a result, GSEs could potentially get higher loans with more confidence and reduced risk, potentially leading to increased loan approvals and more profit for lenders.

Additionally, I advise real estate investors to invest in remodeled, owner-occupied properties with more bathrooms for a higher selling price for a single-family home. This study showed that remodeled properties in metro areas sell faster than non-remodeled properties in rural areas. Investors could potentially buy houses at a lower price and renovate to maximize returns.

In conclusion, GSEs and investors should work together to advance current appraisal methods, which rely on past data to a robust AI automated valuation model (AVMs). By using large scale data in partnership with professional appraisers, they can dynamically adjust valuation based on market conditions and buyers' preference which are updated over time. This approach will improve appraisal accuracy, better align with contract prices and enhance confidence in loan approvals.

# Works Cited


Leonard, D. (2013, Jan 15). *Consumers are Willing to Pay for Quality*. From Custom Builder: https://www.custombuilderonline.com/home/blog/55175184/consumers-are-willing-to-pay-for-quality

Cadena, A., & Thomson, T. A. (2020, June). The Value of "Green" in Resale Residential Real Estate: Premium by Neighborhood Value Quintile, Homestead Status and Year. *Journal of Sustainable Real Estate, 13*(1), 12-29.

Hagen, D. (2005). Real Estate Prices: City Premiums and Neighborhood Effects Real Estate Prices: City Premiums and Neighborhood Effects. *Undergraduate Economic Review Undergraduate Economic Review*.

English, R. (2024). *Unveiling the Dynamics of Home Sale Prices and Time on Market Trends*. From Robbie English: https://www.robbieenglish.com/blog/unveiling-the-dynamics-of-home-sale-prices-and-time-on-market-trends-in-2023/

Laurent Berge. (n.d.). *fixest: Fast and user-friendly fixed-effects estimation.* From lrberge.github.io: https://lrberge.github.io/fixest/

Andrews, e. (2024). *Half of homes sold appraised for more than the sale price*. From housingwire.com: https://www.housingwire.com/articles/half-of-homes-sold-appraised-for-more-than-the-sale-price/

Eriksen, M. D., Fout, H. B., Palim, M., & Rosenblat, E. (2019). *The influence of contract prices and relationships on collateral valuation.* From https://www.fanniemae.com/: https://www.fanniemae.com/media/48121/display

# Appendix

| | Variable 1 | Variable 2 | P-Value | Cramér's V |
|---|---|---|---|---|
| 1 | GR | MT | 0.000 | 0.065 |
| 2 | OO | MT | 0.000 | 0.014 |
| 3 | Pvt | MT | 0.000 | 0.215 |
| 4 | Bup | MT | 0.000 | 0.138 |
| 5 | DS | MT | 0.000 | 0.281 |
| 6 | LS | MT | 0.000 | 0.137 |
| 7 | Q | MT | 0.000 | 0.065 |
| 8 | Con | MT | 0.000 | 0.028 |
| 9 | Bath | MT | 0.000 | 0.019 |
| 10 | Bed | MT | 0.000 | 0.011 |
| 11 | Upd | MT | 0.000 | 0.050 |
| 12 | GLA | MT | 0.000 | 0.039 |
| 13 | DR | MT | 0.000 | 0.216 |

Appendix Table 1: Cramers V table association between Marketing Time and other Categorical Variables

From table 1, DS (Demand and Supply) has moderate association with MT (Marketing Time) at 0.28 showing moderate association between MT and DS. This does not show comprehensive results; therefore regression analysis was done in this study.

## Question 1: R code for applying regression analysis

```
# Question 1: What variables effect the CTA

#regression model with Fixed effects - for hypothesis 1
#Upd, Bath, Bed and Quality as independent variable
model_question_CTA_hypothesis1<- plm(CTA ~  Upd + Bath + Bed + Q,
                                data = data_for_analysis,
                                index = c("Yr", "CFips"),
                                model = "within")

#showing model summary for hypothesis 1
summary(model_question_CTA_hypothesis1)


#regression model with Fixed effects - for hypothesis 2
#GR, Bath, Bed and Quality as independent variable
model_question_CTA_hypothesis2 <- plm(CTA ~  GR + DS + DR + Bed + Bath,
                                data = data_for_analysis,
                                index = c("Yr", "CFips"),
                                model = "within")
#showing model summary for hypothesis 2
summary(model_question_CTA_hypothesis2)


#regression model with Fixed effects - for hypothesis 3
#DR, Lot Size, Bath, Bed and Quality as independent variable
model_question_CTA_hypothesis3 <- plm(CTA ~  DR + LS + Bath + Bed + Q,
                                data = data_for_analysis,
                                index = c("Yr", "CFips"),
                                model = "within")
#showing model summary for hypothesis 3
summary(model_question_CTA_hypothesis3)
```

## Question 2: R code for applying regression analysis

```
# Question 2: What variables effect the marketing time

# Fit the model with fixed effects - Year and County
model_question_hypothesis4 <- feglm(MT ~ CTA + DR + Upd + GR | Yr + CFips,
                                family = binomial(link = "logit"),
                                data = data_for_analysis)

# Summary of the model
summary(model_question_hypothesis4)
```