WORKING PAPER SERIES

Title

Using AI to Improve Price Transparency in Real Estate Valuation

Authors

Cunjia Xu

William C. Wheaton

Date

December 2024

Center for Real Estate

MIT Center for Real Estate Research Paper No.

24/16

Center for Real Estate

Using AI to Improve Price Transparency in Real Estate Valuation
Cunjia Xu, William C. Wheaton
MIT Center for Real Estate Research Paper No. 24/16
December 2024

Abstract

This thesis explores the integration of artificial intelligence (AI) into real estate valuation, focusing on visual property attributes to enhance traditional Hedonic models. By incorporating Vision Language Models (VLMs) and generative AI, the research evaluates the potential of these technologies to assess non-standard variables like aesthetic appeal, condition and cohesiveness of interior and exterior property photos. The study contrasts traditional hedonic regression models, which rely on quantifiable factors such as square footage and location, with a new approach that includes AI-generated scores derived from property photos. The study employs three distinct models: the No_Rubric Model, the Composite Model, and the Verbose Model with the Hedonic model serving as the baseline for evaluating their performance. The results demonstrate that incorporating visual data significantly improves model accuracy, aligning valuations more closely with buyer preferences and sold prices. This shift addresses the industry's need for price transparency and highlights how developers can design properties that better meet market demands.

Cunjia Xu
Massachusetts Institute of Technology

William C. Wheaton
Massachusetts Institute of Technology

Chapter 1

Introduction

1.1 Context and Motivation

Transparency is the cornerstone of any efficient and fair marketplace, and the real estate sector is no exception. When transparency falters, stakeholders encounter significant challenges that undermine trust, fairness, and efficiency. In real estate, price transparency plays a critical role by ensuring that accurate and comprehensive information about property values is available and accessible to all participants—buyers, sellers, and investors alike. By providing all market participants, including buyers, sellers, and investors, with a complete understanding of pricing dynamics, transparency enables informed decision-making. It minimizes the risk of information asymmetry, where one party holds an unfair advantage due to having more information, and promotes a fair and efficient marketplace. Ultimately, transparency contributes to a more competitive environment that supports equitable and data-driven transactions.

A lack of transparency in the real estate market can lead to multiple issues. One of the key problems is information asymmetry, where certain parties, such as sellers or real estate agents, have more information than buyers. Kask and Maani emphasize that this imbalance can distort the perceived value of a property and lead to biased pricing outcomes [1]. They point out that when market participants have unequal access to information, it becomes difficult to make accurate assessments, resulting in "transformation bias" and "information bias," which further contribute to market inefficiencies. Moreover, the absence of transparency can cause broader market inefficiencies. Goodman explains that when participants do not have access to accurate and comprehensive information, it disrupts the price equilibrium, leading to mispricing and reduced market liquidity [2]. Such inefficiencies can create instability within real estate markets, making it difficult for buyers and sellers to make informed decisions.

As illustrated in Fig. 1.1, discussed in Section 1.1, real estate markets around the world ex-

Transparency Level	2024 Composite Rank	Market	2024 Composite Score
High	1	United Kingdom	1.24
	2	France	1.26
	3	United States	1.34
	4	Australia	1.37
	5	Canada	1.49
	6	Netherlands	1.49
	7	New Zealand	1.59
	8	Ireland	1.72
	9	Sweden	1.77
	10	Germany	1.79
	11	Japan	1.83
	12	Belgium	1.84
	13	Singapore	1.92
Transparent	14	Finland	1.97
	15	Hong Kong SAR	1.97
	16	Denmark	2.04
	17	Switzerland	2.05
	18	Spain	2.06
	19	Italy	2.12
	20	Poland	2.13
	21	Norway	2.24
	22	Czech Republic	2.27
	23	Luxembourg	2.29
	24	Hungary	2.30
	25	Portugal	2.30
	26	Chinese Taipei	2.34
	27	South Korea	2.35
	28	UAE - Dubai	2.38
	29	South Africa	2.40
	30	China - Tier 1	2.42
	31	India - Tier 1	2.44
	32	Thailand	2.53
	33	Malaysia	2.57
	34	Romania	2.61
	35	Slovakia	2.62
Semi	36	Greece	2.71
	37	Mexico	2.77
	38	Saudi Arabia	2.79
	39	Israel	2.79
	40	Indonesia	2.81
	41	UAE - Abu Dhabi	2.87
	42	Brazil	2.89
	43	Bulgaria	2.91
	44	Croatia	2.92
	45	Philippines	2.95
	46	Turkey	2.96
	47	Chile	3.06
	48	Peru	3.14
	49	Vietnam	3.25
	50	Slovenia	3.25
Semi	51	Kenya	3.31
	52	Argentina	3.36
	53	Serbia	3.37
	54	Macao SAR	3.42
	55	Colombia	3.46
	56	Mauritius	3.47
	57	Puerto Rico	3.47
Low	58	Malta	3.54
	59	Morocco	3.55
	60	Botswana	3.62
	61	Egypt	3.64
	62	Zambia	3.68
	63	Sri Lanka	3.69
	64	Nigeria	3.69
	65	Bahrain	3.79
	66	Pakistan	3.87
	67	Costa Rica	3.87
	68	Qatar	3.89
	69	Uruguay	4.00
	70	Jordan	4.02
	71	Oman	4.14
	72	Rwanda	4.14
	73	Ghana	4.15
Opaque	74	Ecuador	4.19
	75	Algeria	4.37
	76	Tunisia	4.38
	77	Angola	4.40
	78	Panama	4.40
	79	Uganda	4.40
	80	Mozambique	4.40
	81	Ivory Coast	4.42
	82	Lebanon	4.43
	83	Tanzania	4.44
	84	Senegal	4.47
	85	Honduras	4.46
	86	Dominican Republic	4.48
	87	Guatemala	4.54
	88	Ethiopia	4.57
	89	Iraq	4.60

Figure 1.1: This chart illustrates the composite rankings and scores of global real estate markets across transparency levels, categorized as High, Transparent, Semi-Transparent, Low, and Opaque [3].

hibit varying levels of transparency, ranging from “Highly Transparent” to “Opaque”. Leading countries such as the United Kingdom, United States, France, and Australia rank highest in transparency, benefiting from enhanced technology integration, robust data availability, and progressive climate reporting. These transparent markets have set benchmarks, attracting over 1.2 trillion in commercial real estate investments over the past two years, accounting for more than 80 percent of global activity. Meanwhile, regions with lower transparency, such as certain parts of Asia, Africa, and the Middle East, still hold significant potential for improvement, particularly as urbanization accelerates and demands for infrastructure grow. Real estate transparency is more critical than ever amid global uncertainties, from geopolitical tensions to rapid technological advances. Generative AI (GenAI) is further pushing the boundaries of transparency by automating data processing, property valuations, and urban design analytics, setting the stage for a new era of real estate intelligence. These advancements highlight the competitive edge that transparency brings to market participants, positioning the most transparent markets as leaders in the next phase of global urban growth.

The Hedonic Pricing Model (HPM) is a widely applied approach for estimating property values by examining various characteristics that impact price. It seeks to enhance price transparency by breaking down a property’s price into its key components, such as location, size, age, and amenities. This dissection offers a clearer understanding of the factors that contribute to a property’s value. As outlined by Rosen, goods in differentiated markets, like real estate, are valued based on their utility-bearing attributes, which can be objectively measured [4]. Rosen’s model emphasizes that the observed prices of properties and their associated characteristics establish a set of implicit or “hedonic” prices. These implicit prices help in identifying the key elements that influence property values. Additionally, the HPM provides a holistic view by considering both internal factors, like property features, and external factors, such as neighborhood quality. Rosen’s framework treats differentiated products as a function of their characteristics and examines how equilibrium prices shape the decisions of both consumers and producers.

Despite its advantages, the Hedonic Pricing Model may not capture all factors affecting price if certain information is hidden or not disclosed, such as renovation date. Rosen’s work points out that hedonic prices are determined by the availability of accurate information about the characteristics that differentiate products. When critical characteristics are omitted or obscured, the model’s estimations become less reliable, leading to potential misevaluations [4]. Additionally, some quality aspects of properties, such as visual elements depicted in photos, are challenging to quantify using traditional hedonic models. For instance, visual appeal and the cohesiveness of design require advanced methodologies to analyze effectively. Halvorsen and Pollakowski indicate that many existing models fail to incorporate such non-quantifiable visual factors due to limitations in functional forms and available data [5]. They emphasize that the choice of functional form significantly impacts the model's ability to capture the complex interactions between property characteristics and market value.

1.2 Research Objective

This research aims to develop and demonstrate a methodology for enhancing the traditional Hedonic Regression Model (HRM) by integrating AI-generated scores based on property photos through Vision Language Models (VLMs). This approach aim to systematically incorporate visual property attributes—often omitted in conventional models—into the valuation process, ultimately improving property valuation accuracy and revealing deeper insights into buyer preferences.

Chapter 2

Literature Review

2.1 Hedonic Regression in Real Estate

The Hedonic pricing model, according to Rosen, is a foundational approach in real estate economics that decomposes property prices based on individual attributes like bedroom and bathroom count, square footage, and location. Each characteristic contributes to the overall property price, making it possible to isolate and estimate the value added by each feature. Common functional forms include the linear model, which assumes each additional unit of an attribute (e.g., an extra bedroom) has a constant effect on price, adding a fixed dollar amount regardless of the initial number. The log-linear model instead applies a logarithmic transformation to both independent and dependent variable, allowing for the interpretation of coefficients as percentage changes in price per 1 percent change in each attribute. This model effectively captures exponential relationships, where attributes contribute proportionally rather than absolutely. The semi-log model instead applies a logarithmic transformation to the price, interpreting coefficients as percentage changes; for instance, a bedroom coefficient of 0.05 implies a 5 percent increase in price with an additional bedroom. The translog model further expands by introducing interaction terms, allowing for more complex relationships between attributes, such as how an increase in both bedrooms and lot size might collectively amplify a property's value more than their separate effects would suggest. Each functional form thus provides unique insights into how property characteristics influence market value, with transformation choices reflecting the expected relationships between features and price [5].

2.2 AI Applications in Real Estate Pricing

In recent years, the integration of artificial intelligence (AI) in real estate pricing has gained significant attention, with research exploring various AI-driven methods to enhance property valuation accuracy. Traditional hedonic pricing models struggle to incorporate qualitative, visual. AI-based approaches, including Computer Vision (CV), Machine Learning (ML), and advanced neural network architectures, have been introduced to bridge this gap, allowing for a more comprehensive valuation process. By leveraging CV to interpret property images and ML algorithms to capture complex, non-linear relationships between features, these AI methods add an essential layer of quantitative assessment on visual factors that are hard to quantify previously.

CV, particularly through Convolutional Neural Networks (CNNs), provides a systematic approach to analyzing property images, converting these qualitative characteristics into measurable attributes that complement hedonic data.

In a study by Poursaeed, CNNs were utilized to analyze real estate images to estimate the luxury level and condition of properties [6]. The researchers compiled an extensive dataset of over 200,000 images, featuring both interior and exterior views from various rooms and angles per property. To ensure image consistency, all photos were standardized in terms of size, brightness, color, and contrast, resulting in a uniform dataset suitable for model analysis. Each image was then labeled with luxury levels, such as “high-end” or “basic,” through a crowdsourcing framework, which enabled the CNN to learn and recognize visual patterns associated with luxury indicators, like high-quality finishes, spacious layouts, and distinct architectural details. During training, the model categorized each image into one of eight luxury levels. Images with features that closely resembled high-luxury examples received higher scores, showing a ranking based on their similarity to luxury features. In terms of performance, the model achieved a 5.8 percent median error rate compared to Zillow’s 7.9 percent.

In You’s study, deep learning was leveraged, particularly convolutional and recurrent neural networks, to enhance property valuation by integrating visual content from property images and spatial context from neighborhood data [7]. CNNs are used to extract visual features from property images, capturing elements like curb appeal and design aesthetics, while Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, model the relationships between properties within the same neighborhood. This combination allows the model to incorporate both visual appeal and location-based price influences, offering a more comprehensive approach to real estate appraisal. The model performed best in data-rich environments, achieving high accuracy in San Jose with a Mean

Absolute Error (MAE) of 17.98 and Mean Absolute Percentage Error (MAPE) of 4.58 percent, outperforming traditional appraisal methods and baseline models.

Vision Transformers (ViTs) also represent a notable advancement in computer vision. Unlike traditional CNNs that process images by focusing on localized features, ViTs employ a self-attention mechanism, allowing them to learn and capture relationships across an entire image [8]. ViTs divide an image into fixed-size patches, which are then embedded as vectors and fed into a Transformer model, enabling a global understanding of the image.

In the study by Yazdani and Raissi, the ViT model captures specific visual attributes from property images, such as the style of interior design (e.g., modern, traditional), the quality and condition of finishes (e.g., new, renovated, worn), the presence of architectural elements (e.g., fireplaces, high ceilings, large windows), and exterior features like landscaping or curb appeal [9]. These nuanced visual aspects, which are often influential in buyer preferences, are distilled into numerical feature vectors—quantified representations that encapsulate each property's visual characteristics. The visual characteristics captured by the ViT model are combined with quantitative data, and neighborhood factors. This enriched dataset is then used in a regression model to predict property values. The results show that this approach improves prediction accuracy, with a 10.63 percent reduction in Root Mean Squared Error (RMSE) compared to a baseline model that only used quantitative features.

While CV models like CNNs and ViTs analyze the visual quality of images, Machine Learning (ML) algorithms such as Random Forests and Support Vector Machines (SVMs) are often employed to integrate these visual scores with traditional hedonic data. Unlike CV models, ML algorithms are not directly used to assess image quality but instead focus on optimizing the combination of visual scores with traditional numeric data. ML models help identify non-linear relationships that are beyond the scope of hedonic models. In Kintzel's study, for example, use machine learning approaches, exploring neural networks, random forest, and gradient boosting models, to enhance house price prediction models [10]. "Our best model created predictions of house price that were accurate within 10 percent of the actual sale price over 63 percent of the time".

These advancements in computer vision and machine learning underscore the transformative potential of AI in property valuation, setting a strong foundation for refinement of traditional pricing models.

Chapter 3

Methodology

3.1 Data Source and Preparation

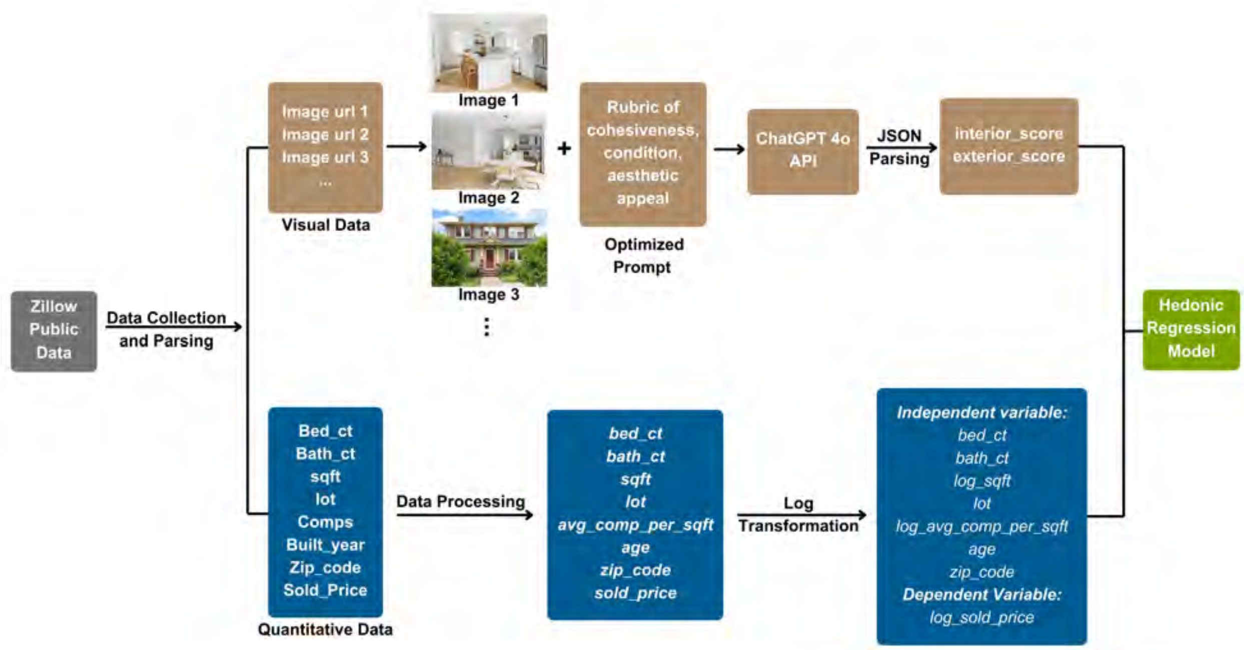


Figure 3.1: The workflow diagram demonstrates how visual and quantitative data are processed through a rubric, ChatGPT API, and log transformations to generate inputs for a hedonic regression model for property valuation.

This study integrates both qualitative image data and quantitative features to enhance the hedonic regression model for predicting the sales prices of single-family homes. As shown in Fig. 3.1, the quantitative data is sourced from Zillow's public website and includes key property attributes such as the number of bedrooms (bed_ct), number of bathrooms (bath_ct), square footage (sqft), lot size (lot), comparable prices (comps), and the year the property was built (built_year). Zillow provides 3 to 5 comparable prices for each property, representing sale prices of similar nearby properties sold close to the property's sale date, thereby giving a more accurate context for each property's estimated value. Each property is also identified by a unique Zillow property ID (zpid), enabling precise tracking of individual properties. To control for location, zip codes are included as dummy variables, while property sale dates are restricted to those between January 1 and December 31, 2023, to reduce the influence of temporal factors. This study focuses on properties in Arlington, MA, and Woburn, MA, allowing for a controlled comparison between two specific market areas.

The qualitative component consists of property photos, also sourced from Zillow. Each property has at least one interior and one exterior photo, with all available images included in the analysis. These images provide additional context, capturing visual qualities that may impact buyer perception and, consequently, property prices.

(a) Exterior View

(b) Kitchen

(c) Dinning Room

(d) Living Room

(e) Bathroom

(f) Patio

Figure 3.2: Sample Images from a Property Listing in Dataset [11].

3.2 Data Processing

During the data processing phase, as depicted in Fig. 3.1, the raw dataset of 310 properties—150 from Arlington, MA, and 160 from Woburn, MA—underwent several transformations to ensure accuracy and consistency. First, each property's built year was converted to property age (age) by subtracting the built year from the current year, 2024. For comparable prices (comp_price) provided by Zillow, values were adjusted to a per-square-foot basis to standardize across varying property sizes, with an average calculated across the 3 to 5 comparable properties listed for each home (avg_comp_per_sqft).

Data collection, parsing, and cleaning were conducted using Python, prioritizing completeness and accuracy. Properties missing essential quantitative variables—such as bedroom count, bathroom count, or comparable prices—or those lacking images were removed, reducing the dataset by 13 properties. An additional 30 properties were excluded due to incomplete visual data (e.g., containing only interior or only exterior photos). This filtering process resulted in a final dataset of 267 properties.

To further prepare the data for modeling, several variables were log-transformed to capture percentage changes rather than absolute unit changes, addressing potential non-linear relationships with the dependent variable. Specifically, the dependent variable (log_sold_price) was transformed to its logarithmic form. Additionally, square footage (log_sqft) and average comparable price per sqft (log_avg_comp_price_per_sqft) were log-transformed, as these variables often exhibit diminishing marginal effects on property value, making the log transformation suitable for linearizing such relationships and improving model interpretability.

The processed dataset was then divided into training and testing sets, with 90 percent of the data used to train the model and the remaining 10 percent reserved for testing. This split enables the evaluation of model performance on unseen data, ensuring the robustness and predictive accuracy of the final hedonic regression model.

3.3 AI-Generated Property Photo Scores

To incorporate qualitative visual data into the hedonic model, a Vision Language Model (VLM) using the pre-trained ChatGPT-4o API was selected to analyze and score property photos, both interior and exterior. As illustrated in Fig. 3.1, this process involves sourcing image data from Zillow and applying a structured scoring rubric to capture key qualitative features, focusing on aesthetic appeal, condition, and design cohesiveness. Each property includes multiple images, with a minimum of one interior and one exterior photo; all available images for each property were analyzed to ensure comprehensive visual assessment.

For each qualitative dimension, a defined rubric and prompt were provided to the ChatGPT-4o model to standardize the evaluation process and maintain consistency across properties. These prompts instructed the model to assess each property image on a scale of 1 to 5 for aesthetic appeal (capturing the visual attractiveness and style), condition (indicating the maintenance level and quality of finishes), and design cohesiveness (measuring how well elements fit together visually).

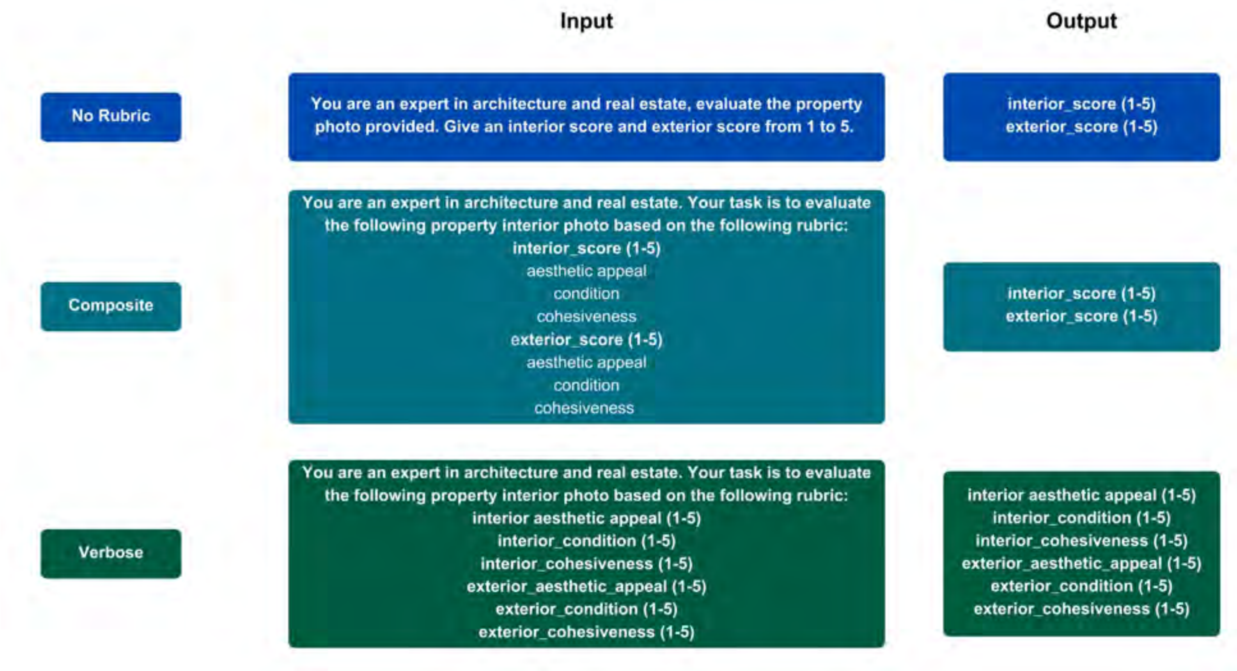


Figure 3.3: No Rubric vs. Composite vs. Verbose. The figure illustrates three distinct scoring approaches—No Rubric, Composite, and Verbose—used to evaluate property attributes via ChatGPT-4o. Each method differs in the level of detail provided for input criteria and output scores, ranging from general scores (No Rubric) to detailed breakdowns of visual characteristics (Verbose).

Three distinct scoring methods were used to capture visual attributes:

No_Rubric Method: In this approach, ChatGPT-4o was asked to score property attributes without any specific guidelines or criteria. This means the model is prompted to provide general scores on the interior and exterior photos.

Composite Method: In this approach, ChatGPT-4o was prompted to provide two overall scores for each property—one for the interior and one for the exterior. Each score reflects an aggregated assessment across the three dimensions (aesthetic appeal, condition, and design cohesiveness), resulting in a single composite score for the interior and another for the exterior.

Verbose Method: In this approach, ChatGPT-4o assigned six scores per property, evaluating aesthetic appeal, condition, and design cohesiveness separately for interior and exterior photos. This resulted in three distinct scores for each category (interior and exterior), offering a detailed breakdown of the property's visual characteristics.

The motivation for testing different prompt formats arises from the need to evaluate whether a structured rubric is critical for enhancing model performance. Furthermore, examining different levels of scoring granularity helps assess the degree of nuance required in scoring criteria to accurately capture key visual attributes and optimize model outcomes.

Figure 3.4: AI Score Generating Workflow. The diagram depicts the AI score-generating workflow, where interior and exterior property images, combined with a prompt, are processed via the ChatGPT-4o API in a single session.

As shown in Fig. 3.4, all images—including the interior and exterior photos—along with the prompt are fed into the GPT API in a single API call. This process yields both the interior and exterior scores simultaneously for the composite model and provides six individual criteria scores through the verbose method.

Below is an example of the interior rubric and prompt used for composite method; the full rubric and prompt for all methods are available in the Appendix.

Prompt: *You are an expert in architecture and real estate. Your task is to evaluate the following property interior photo based on cohesiveness, condition and aesthetic appeal. Use the following rubric:*

Score: 1 (Poor) Cohesiveness: The design is disjointed, with clashing architectural styles or materials that create a chaotic, incoherent appearance. Condition: The property shows significant visible damage, neglect, or deterioration, such as peeling paint, broken windows, or structural issues. Trash, debris, or overgrown vegetation are present, indicating poor cleanliness. The landscaping appears dead or poorly maintained, with no evident effort in upkeep. Overall, the property looks run-down and requires major repairs to restore its condition. Aesthetic Appeal: The property photo fails to capture any appealing elements. The composition is poor, and the image looks dull, with unappealing angles or lighting. There is little to no effort in showcasing the property's potential charm. The photo makes the property look uninviting or unattractive, and the exterior appears neglected or out of place in its surroundings.

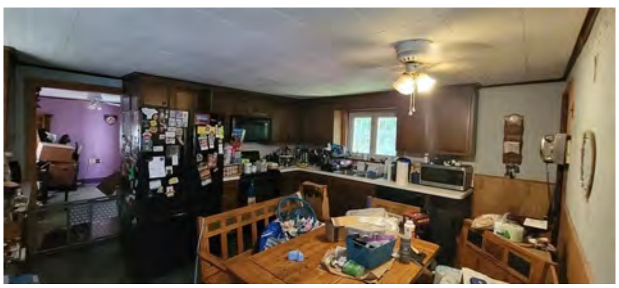
Score: 2 (Fair) Cohesiveness: Some stylistic inconsistencies exist, but the overall design still feels somewhat connected, though imperfect. Condition: The property has noticeable *issues like chipped paint, minor cracks, or missing roofing materials. Although wear and tear are visible, there are no immediate structural concerns. The cleanliness is below average, with some debris or clutter in the yard, and yard maintenance appears minimal. Landscaping exists but is overgrown or lacks proper care. The property looks functional but worn, with small updates necessary to improve its condition. Aesthetic Appeal: The property photo shows some effort but lacks visual appeal. The angle or lighting is not ideal, resulting in a somewhat flat or uninspired image. The property's attractive features are either not emphasized or are overshadowed by less appealing elements. While the photo provides a basic view of the property, it does not highlight its full potential or charm, leaving much to be desired in terms of aesthetics.*

Score: 3 (Average) Cohesiveness: The architectural style is consistent, but lacks strong integration or thematic unity. The design is cohesive but unremarkable. Condition: The property is generally well-maintained, with minor cosmetic issues such as slight discoloration or light wear and tear. There is no major visible damage, and the yard is clean with minimal debris. Landscaping is maintained but could benefit from further attention or updates, such as trimming or new plantings. The property has a solid visual appeal but may not stand out as exceptional. Aesthetic Appeal: The property photo is decent, showing the exterior in a fairly positive light. The angles and lighting are reasonable, offering a clear view of the property without any major flaws in composition. The photo captures some attractive elements of the property but lacks any standout features that make it truly eye-catching. Overall, the property looks appealing but not exceptional.

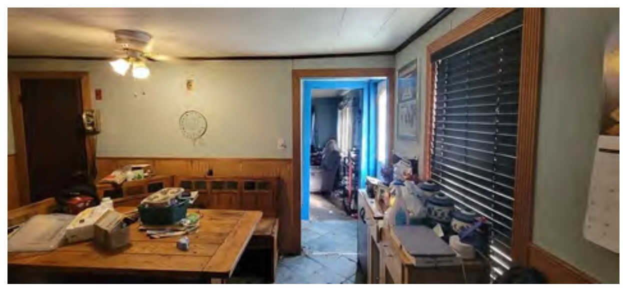
Score: 4 (Good) Cohesiveness: The design shows a clear and consistent architectural style with well-coordinated materials and elements that complement each other. Condition: The property is well-maintained with only minor, easily fixable issues like slight fading or minimal wear. Cleanliness is high, with no visible clutter or debris, and the yard appears well-kept. Thoughtful landscaping adds to the property's appeal, with manicured lawns, pruned trees, and healthy plants. The overall appearance is attractive and reflects good care, with only minor enhancements needed for perfection. Aesthetic Appeal: The property photo is well-composed, with good angles and lighting that enhance the property's exterior appeal. The image captures key attractive features, such as architectural details or landscaping, creating a visually appealing photo. There is a clear effort to make the property look inviting and charming, and the overall aesthetics are strong

Score: 5 (Excellent) Cohesiveness: The design is perfectly cohesive, with every architectural element working together seamlessly to form a unified, harmonious structure. Condition: The property is in pristine condition, with no visible issues. It appears new or recently renovated, showing no signs of damage or wear. The cleanliness is impeccable, with no clutter or neglect *in sight. High-quality landscaping, such as sculpted trees, fresh flowers, and lush greenery, adds significant curb appeal. Overall, the property exudes high value and exceptional care, looking immaculate and well-preserved. Aesthetic Appeal: The property photo is visually stunning and expertly composed. The lighting, angles, and focus are perfect, highlighting the most attractive aspects of the property and making it look incredibly inviting. Architectural details, landscaping, and surroundings are captured in a way that enhances the property's beauty, creating a picture-perfect image. The photo exudes a sense of luxury or charm, making the property stand out as highly desirable and aesthetically pleasing.*

Here are examples of properties that received the highest and lowest scores across each dimension for both interior and exterior images, illustrating how ChatGPT-4o's VLM assesses and interprets their visual characteristics.

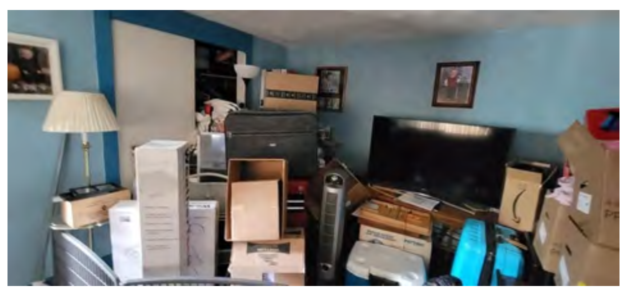


(a) Kitchen



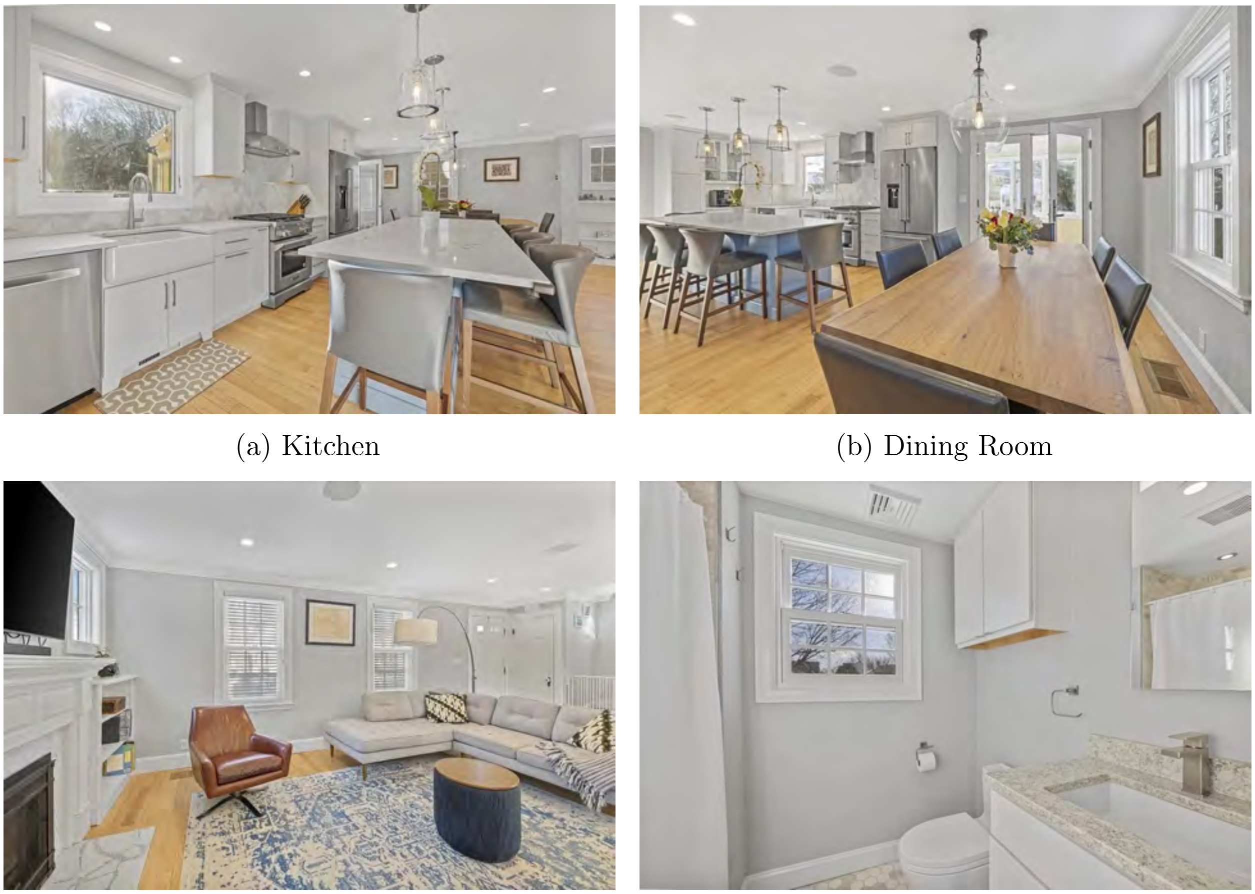
(b) Dining Room

(c) Living Room



(d) Living Room

Figure 3.5: Interior Cohesiveness: Score 1 (Poor)
AI score explanation: The interior of the property presents a cluttered and chaotic appearance with various mismatched design styles and colors.



(a) Kitchen

(b) Dining Room

(c) Living Room

(d) Bathroom

Figure 3.6: Interior Cohesiveness: Score 5 (Excellent)
AI score explanation: The interior of the property demonstrates a clear and consistent design style, utilizing modern elements and neutral tones that complement each other well.

(a) Kitchen

(b) Dining Room

(c) Living Room

(d) Bathroom

Figure 3.7: Interior Aesthetic Appeal: Score 1 (Poor)
AI score explanation: Aesthetically, the space does not provide striking visuals and lacks balance and refinement.

(a) Kitchen

(b) Dining Room

(c) Living Room

(d) Bathroom

Figure 3.8: Interior Aesthetic Appeal: Score 5 (Excellent)
AI score explanation: The aesthetic appeal is visually stunning, with attention to detail and a sophisticated environment.

(a) Kitchen

(b) Dining Room

(c) Living Room

(d) Bathroom

Figure 3.9: Interior Condition: Score 1 (Poor)
AI score explanation: While there are no major structural issues visible, the overall feel is underwhelming and suggests the need for cosmetic improvements and updates.

(a) Kitchen

(b) Dining Room

(c) Living Room

(d) Bathroom

Figure 3.10: Interior Condition: Score 5 (Excellent)
AI score explanation: The condition of the property is immaculate, with no visible signs of wear or damage; everything appears new or recently renovated, and the space is spotless and well-maintained.

(a)

(b)

Figure 3.11: Exterior Cohesiveness: (a) Score 2 (Fair). AI score explanation: The property has a consistent architectural style, though it is unremarkable. (b) Score 5 (Excellent). AI score explanation: The exterior property images display an exceptionally well-maintained home with a cohesive architectural design. The brick façade, roof tiles, and white trim work together seamlessly, creating a harmonious appearance.

(a)

(b)

Figure 3.12: Exterior Aesthetic Appeal: (a) Score 2 (Fair). AI score explanation: The overall aesthetic appeal is dull, as the images fail to capture attractive elements, leaving a flat and uninspired impression. (b) Score 5 (Excellent). AI score explanation: The photographs are well-composed, with good angles and lighting that enhance the property's exterior appeal.

(a)

(b)

Figure 3.13: Exterior Condition: (a) Score 2 (Fair). AI score explanation: The condition is fair, showing some wear and tear, overgrown vegetation, and minor visible issues. (b) Score 5 (Excellent). AI score explanation: The condition is excellent, with no visible issues, clean landscaping, and an overall pristine appearance, thus scoring a perfect 5.

Chapter 4

Results and Analysis

This section compares the performance of the 3 new models, no rubric method, composite method and verbose method with the base Hedonic Model. The performance of each model is evaluated using R^2, Adjusted R^2, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) on both training dataset and testing dataset. This comparison allows us to determine the impact of adding qualitative image-based scores on the model's predictive accuracy. Additionally, an AI robustness test was conducted to assess the consistency of ChatGPT-4o's scoring across different criterions when evaluating property photos. This test ensures that the AI provides reliable scores for the same property characteristics under repeated assessments.

4.1 Analysis of Visual Information Representation

Table 4.1: Descriptive Statistics of Base Hedonic

Variable	Coefficient	Std_Error	T_Stat	P_Value
const	7.613	0.815	9.342	4.71E-18
bed_count	0.018	0.017	1.045	0.297
bath_count	0.069	0.018	3.803	0.000179
log_sqft	0.496	0.052	9.609	7.08E-19
lot	1.76E-07	3.90E-07	0.452	0.651
log_avg_comp_value_per_sqft	0.388	0.094	4.134	4.83E-05
age	-0.001	0.0003	-2.785	0.005744
zip_1801	-0.379	0.163	-2.326	0.021
zip_2474	-0.139	0.159	-0.878	0.381
zip_2476	-0.080	0.159	-0.501	0.616
Statistics				**Value**
R_squared				0.827
Adj_R_squared				0.821
Sample_Size				267

Table 4.2: Descriptive Statistic for No Rubric Model

Variable	Coefficient	Std_Error	T_Stat	P_Value
Intercept	8.447	0.694	12.165	0.000
bed_count	0.031	0.015	2.089	0.038
bath_count	0.055	0.015	3.599	0.000
log_sqft	0.373	0.045	8.236	0.000
lot	0.000	0.000	0.641	0.522
log_avg_comp_value_per_sqft	0.275	0.080	3.433	0.001
age	-0.001	0.000	-4.026	0.000
interior_score	0.093	0.015	5.994	0.000
exterior_score	0.072	0.015	4.926	0.000
zip_code_1801	-0.286	0.138	-2.068	0.040
zip_code_2474	-0.036	0.135	-0.271	0.787
zip_code_2476	0.004	0.135	0.027	0.978
Statistics				**Value**
R_squared				0.877
Adj_R_squared				0.872
Sample_Size				267

Table 4.3: Descriptive Statistic for Composite Method

Variable	Coefficient	Std_Error	T_Stat	P_Value
Intercept	8.120	0.673	12.074	0.000
bed_count	0.031	0.014	2.172	0.031
bath_count	0.052	0.015	3.443	0.001
log_sqft	0.368	0.044	8.361	0.000
lot	0.000	0.000	0.425	0.671
log_avg_comp_values_per_sqft	0.356	0.077	4.612	0.000
age	-0.001	0.000	-2.448	0.015
interior_score	0.088	0.014	6.429	0.000
exterior_score	0.071	0.017	4.203	0.000
zip_code_1801	-0.337	0.134	-2.511	0.013
zip_code_2474	-0.133	0.131	-1.020	0.309
zip_code_2476	-0.097	0.131	-0.743	0.458
Statistics				**Value**
R_squared				0.884
Adj_R_squared				0.879
Sample_Size				267

In the Composite regression model shown in Table 4.3, the coefficients for the interior score (β_{interior}) and exterior score (β_{exterior}) are 0.088 and 0.071, respectively, indicating their impact on the natural logarithm of the sold price (log(sold_price)). These coefficients can be interpreted as the percentage change in the sold price for a one-unit increase in the respective score, holding all other variables constant.

The percentage change in the sold price due to a one-unit increase in the interior or exterior score is calculated as:

$$\Delta P = (\exp(\beta \cdot 1) - 1) \times 100\%,$$

where β represents the coefficient of the respective variable. Specifically, for the interior score, the sold price increases by 8.8% ($\exp(0.088) - 1 = 0.088$), and for the exterior score, the sold price increases by 7.1% ($\exp(0.071) - 1 = 0.071$).

To interpret the effects of these scores across their full range (from 1 to 5), the percentage change in the sold price is calculated as:

$$\Delta P = (\exp(\beta \cdot \Delta\text{score}) - 1) \times 100\%,$$

where Δscore represents the total change in the score. For a four-point increase ($\Delta\text{score} = 4$), the percentage changes are as follows:

$$\Delta P_{\text{interior}} = (\exp(0.088 \cdot 4) - 1) \times 100 = 41.29\%,$$

$$\Delta P_{\text{exterior}} = (\exp(0.071 \cdot 4) - 1) \times 100 = 32.04\%.$$

When both interior and exterior scores increase simultaneously from 1 to 5, the combined percentage change is computed as:

$$\Delta P_{\text{combined}} = (\exp\left((\beta_{\text{interior}} + \beta_{\text{exterior}}) \cdot \Delta\text{score}\right) - 1) \times 100\%,$$

$$\Delta P_{\text{combined}} = (\exp\left((0.088 + 0.071) \cdot 4\right) - 1) \times 100 = 85.44\%.$$

These results highlight the significance of interior and exterior attributes in property valuation, suggesting that properties with higher scores in these dimensions command substantially higher prices.

Table 4.4: Descriptive Statistic for Verbose Model

Variable	**Coefficient**	**Std_Error**	**T_Stat**	**P_Value**
Intercept	7.837	0.699	11.209	0.000
bed_count	0.032	0.015	2.179	0.030
bath_count	0.044	0.016	2.845	0.005
log_sqft	0.411	0.045	9.136	0.000
lot	0.000	0.000	0.683	0.495
log_avg_comp_value_per_sqft	0.339	0.080	4.217	0.000
age	-0.001	0.000	-2.974	0.003
interior_cohesiveness_score	0.066	0.041	1.610	0.109
interior_condition_score	0.053	0.018	2.927	0.004
interior_aesthetic_appeal_score	-0.042	0.045	-0.939	0.348
exterior_cohesiveness_score	0.051	0.024	2.159	0.032
exterior_condition_score	0.039	0.022	1.819	0.070
exterior_aesthetic_appeal_score	0.003	0.021	0.167	0.868
zip_code_1801	-0.313	0.139	-2.256	0.025
zip_code_2474	-0.084	0.135	-0.629	0.532
zip_code_2476	-0.040	0.135	-0.294	0.769
Statistics				**Value**
R_squared				0.880
Adj_R_squared				0.873
Sample_Size				267

The regression results from the four models — Base Hedonic, No Rubric, Composite, and Verbose — reveal the impact of integrating qualitative, image-based scores on model explanatory level. The Base Hedonic Model, with an R-squared of 0.827 and serves as the benchmark for comparing the other models. No Rubric Model increases the R-squared to 0.877 compared to the Composite Model has the R-squared to 0.883 and the Verbose Model achieves an R-squared of 0.879. These results illustrate that adding visual quality assessments consistently improves the model's performance.

Examining the p-values across the four models highlights the significance of individual variables, especially when qualitative image-based scores are added. In the Base Hedonic Model, traditional variables like square footage (log_sqft) and comparable value per square foot (log_avg_comp_value_per_sqft) are highly significant ($p < 0.001$), while some variables, such as lot size, bed_count, and specific zip codes, lack significance ($p > 0.1$). When qualitative scores are introduced in the No_Rubric, Composite, and Verbose Models, the p-values for image-based variables such as interior and exterior scores are mostly significant, indicating that these visual elements contribute valuable information to the model. However, introducing multiple scoring criteria in the Verbose Model results in mixed p-values among criteria; for example, 3 scores remain significant, while others show higher p-values. This suggests that a further correlation matrix analysis may be needed to explore potential multicollinearity.

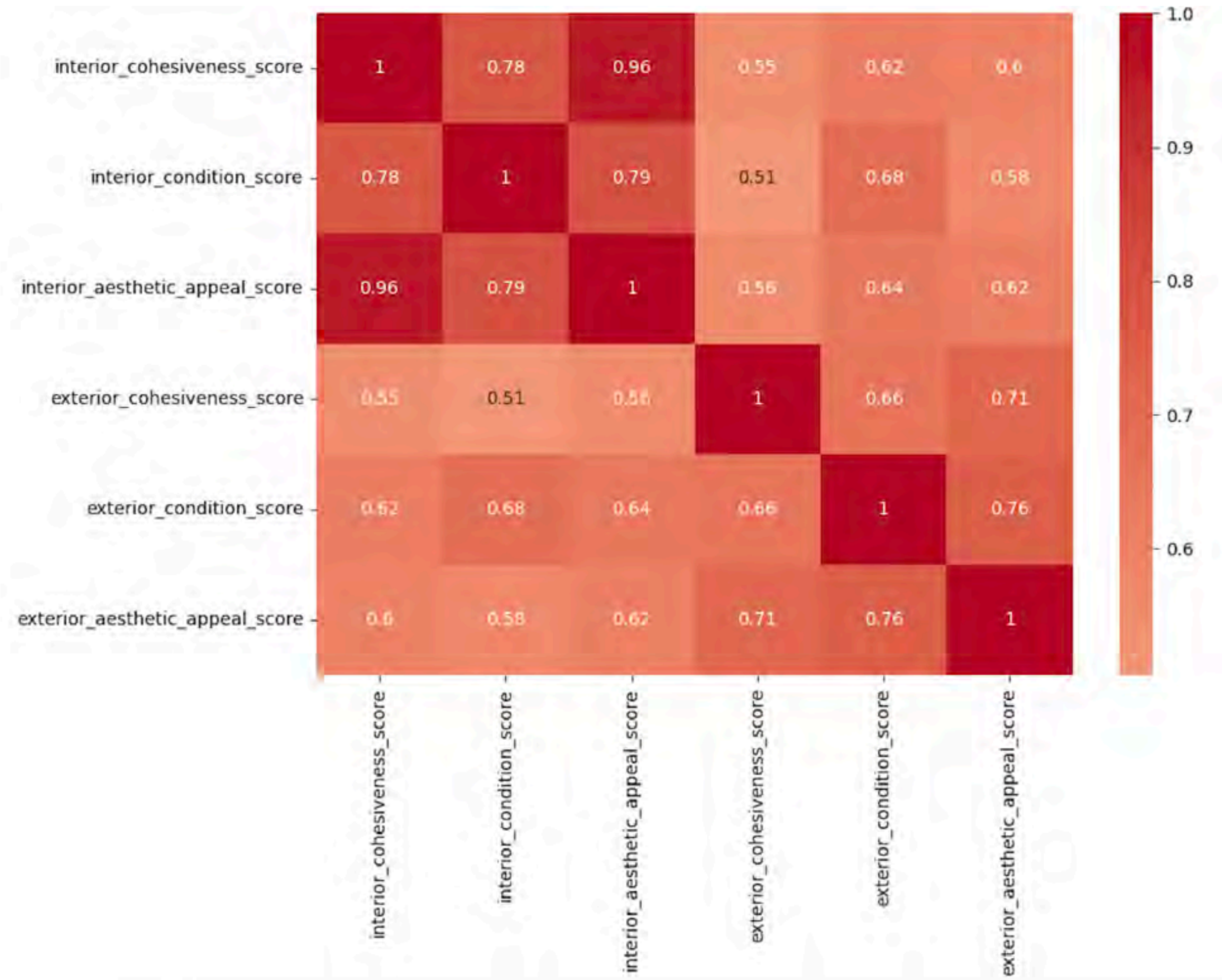


Figure 4.1: Correlation Matrix of Verbose Model

The correlation matrix for the Verbose Model reveals strong relationships among the interior scores, with a particularly high correlation of 0.96 between cohesiveness and aesthetic appeal. Exterior scores also show moderate correlations, while the correlations between interior and exterior scores range from 0.5 to 0.65, indicating weaker relationships.

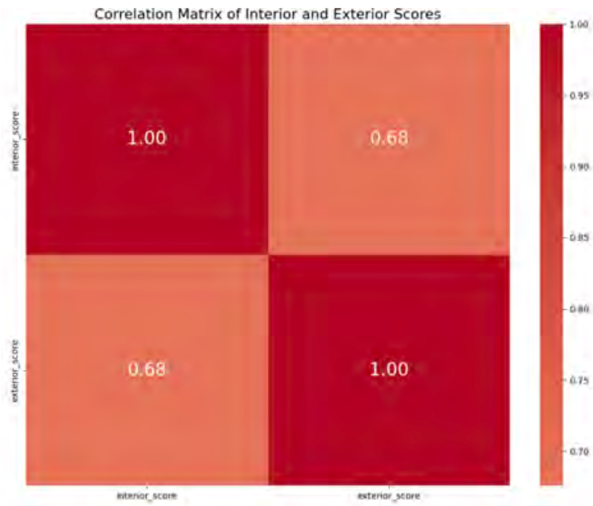


Figure 4.2: Correlation Matrix of Composite Model
The correlation matrix for the Composite Model indicates a single correlation value of 0.68 between the interior and exterior scores. This moderate correlation suggests a balanced relationship between these aggregated dimensions.

The Verbose Model correlation matrix shows that interior scores—cohesiveness, condition, and aesthetic appeal—are highly correlated, with correlation of 0.96 between cohesiveness and aesthetic appeal, 0.79 between aesthetic appeal and condition and 0.78 between condition and cohesiveness. Exterior scores exhibit correlations of 0.66 between cohesiveness and aesthetic appeal, 0.76 between aesthetic appeal and condition and 0.66 between condition and cohesiveness. In contrast, correlations between interior and exterior scores are generally weaker, between 0.5-0.65.

On the other hand, the Composite model shows correlation of 0.68 between interior and exterior score.

While the Verbose Model offers slightly higher explanatory power, it includes overlapping variables that could introduce redundancy and multicollinearity, potentially reducing interpretability. The Composite Model balances explanatory power with a more parsimonious and interpretable set of variables.

For each of the four models, performance metrics, including MAE, RMSE, R^2, adjusted R^2, and MAPE, were calculated for both the training and testing sets.

To ensure robust and reliable results, these metrics were calculated 10 times using 10 different random training and testing splits, yielding an average MAE, RMSE, R^2, adjusted R^2, and MAPE for each model. This approach aims to identify the model with the best overall performance, shown in Table 4.5 and Fig. 4.3.

Table 4.5: Model Comparison Statistic

Metric	**Base Hedonic**	**No Rubric**	**Composite**	**Verbose**
Train R2	0.834 ± 0.008	0.866 ± 0.011	0.863 ± 0.006	0.887 ± 0.013
Test R2	0.785 ± 0.098	0.807 ± 0.102	0.864 ± 0.051	0.811 ± 0.112
Train Adjusted_R2	0.827 ± 0.008	0.859 ± 0.012	0.856 ± 0.007	0.858 ± 0.014
Test Adjusted_R2	0.672 ± 0.150	0.666 ± 0.176	0.765 ± 0.088	0.554 ± 0.265
Train MAE	117607.761 ± 3218.046	102009.055 ± 1723.740	102941.337 ± 3046.143	100566.811 ± 3153.561
Test MAE	117885.657 ± 31147.843	117955.593 ± 17540.469	104372.431 ± 28548.807	117563.340 ± 28975.650
Train RMSE	164403.329 ± 6331.939	144833.377 ± 6365.959	148387.577 ± 5173.360	145072.104 ± 7135.561
Test RMSE	160331.097 ± 52256.627	180183.746 ± 38319.189	138677.696 ± 45770.699	166324.999 ± 55091.715
Train MAPE	12.149 ± 0.308	10.242 ± 0.195	10.146 ± 0.189	9.981 ± 0.246
Test MAPE	12.458 ± 3.037	11.507 ± 1.546	10.637 ± 1.743	11.770 ± 2.165

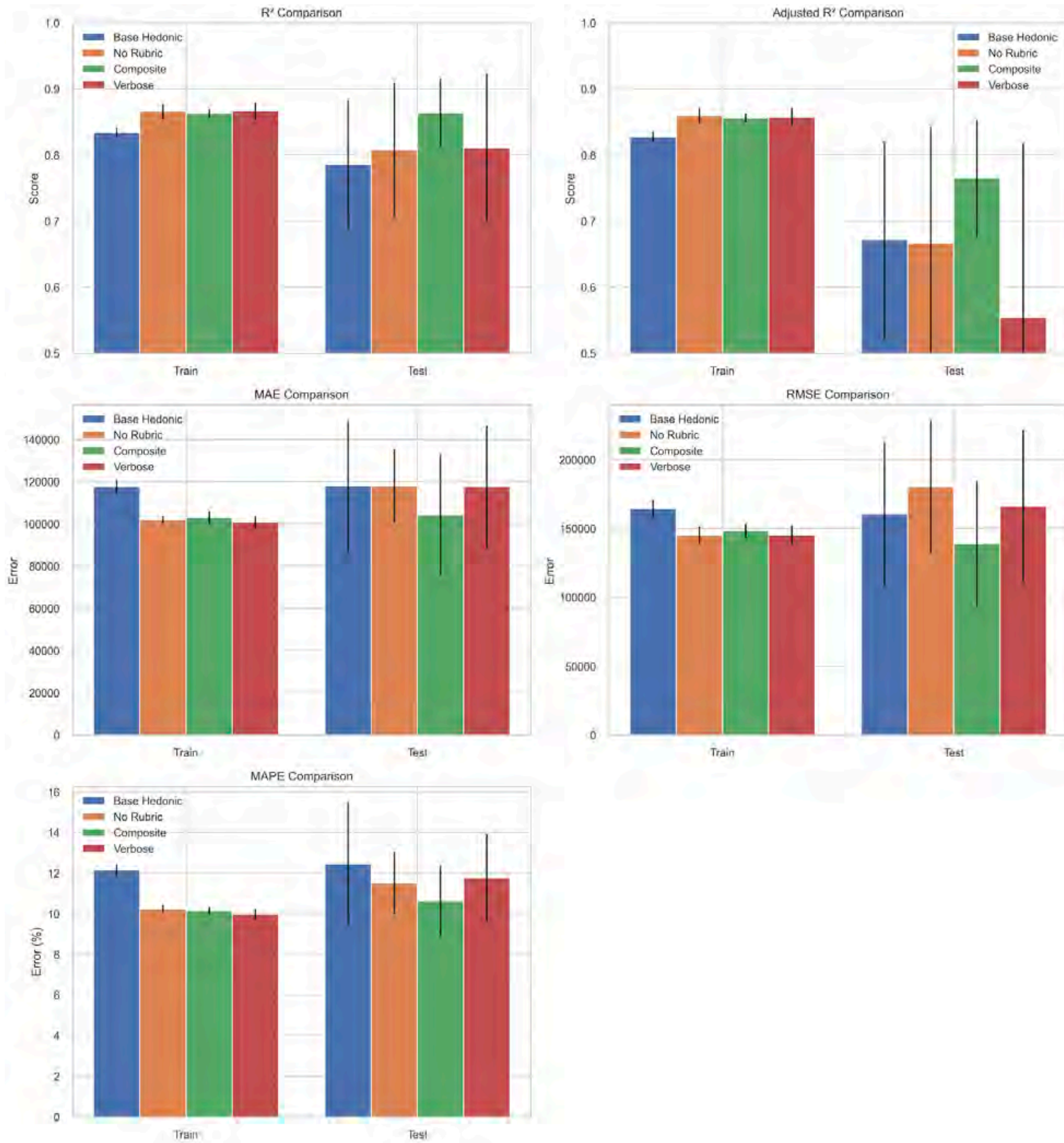


Figure 4.3: Model Comparison Diagram

The No-Rubric model shows a higher R2 on the training dataset, achieving 0.8659 compared to the Base Hedonic model's 0.8338, indicating an improvement in explaining data variance. However, its Adjusted R2 on test dataset is only 0.6662, significantly lower than the Composite model's 0.7646 and close to the Base Hedonic model's 0.6715. This suggests that while the No-Rubric model captures general patterns well, the lack of a specific rubric results in inconsistencies, which is reflected in the high standard deviation of 0.1764 for testing

Adjusted R2 . This variability likely stems from the absence of a clear scoring guideline, causing ChatGPT's scoring for property photos to fluctuate. Consequently, the model's prediction errors are larger, with a Train MAE (Mean Absolute Error) on testing dataset of 102009, higher than those of the Composite model of 104372.4305. Additionally, the Test RMSE (Root Mean Squared Error) is 180183.7461, surpassing the Base Hedonic model's 160331.097, which indicates less stability in specific predictions, especially for extreme values.

The Verbose model achieves a training R2 of 0.8865, which is higher than the Base Hedonic model's 0.8338. However, its Adjusted R2 on the test data is only 0.554, suggesting redundancy among variables, as indicated by a high correlation among them. The high standard deviation of 0.2646 for Test Adjusted R2 and RMSE highlight that the model's predictions vary widely, depending on the specific data segments. This fluctuation in performance implies that the Verbose model's accuracy is inconsistent, with predictions highly dependent on the dataset split.

Finally, the Composite model outperforms the other models in most metrics, achieving the highest R2 on both the training (0.8628) and test datasets (0.8642). It also has the best Adjusted R2 values for both training (0.8562) and test data (0.7646), indicating robust explanatory power without overfitting. The Composite model's well-defined rubric provides consistency in scoring, leading to a lower standard deviation in testing Adjusted R2. It also shows the lowest errors, with a Test MAE of 104372.4305 and a Test RMSE of 138677.6956, indicating both high accuracy and reliability. The RMSE, along with a Test MAPE of 10.6365%, demonstrate that the Composite model not only performs well but also does consistently, establishing it as the most effective model in this comparison.

The Composite model stands out as the best performer because it strikes the optimal balance in the level of detail provided in the prompt. When there is no rubric, the AI scoring becomes inconsistent due to a lack of clear guidance, leading to variability in predictions. Conversely, when the rubric criteria are overly nuanced, as in the Verbose model, the system tends to overfit, capturing redundant or less relevant details. The Composite model effectively avoids these extremes by using a structured yet not overly detailed rubric, providing consistent and meaningful scores that enhance predictive reliability without sacrificing flexibility.

4.2 Robustness Test

To assess the robustness of the AI-generated scores, an additional test was conducted on 9 randomly selected properties. For each property, ChatGPT-4o was used to assign scores for both interior and exterior photos using No_Rubric method and Composite method. This scoring process was repeated 50 times for each property, and the average score and standard deviation were calculated to evaluate the consistency of the AI's scoring for each dimension when using the same prompt. Table 4.6 below shows the results using No_Rubric method while Table 4.7 shows the results using Composite method, illustrating the frequency distribution of scores received across the 50 trials for each dimension.

Table 4.6: No_Rubric Model Score Mean & Std

zpid	dimension	score_1	score_2	score_3	score_4	score_5	mean_score	std_score
56403089	interior_score	0	0	0	2	48	4.960	0.196
56403089	exterior_score	0	0	0	16	34	4.680	0.466
56400471	interior_score	0	0	0	0	50	4.980	0.140
56400471	exterior_score	0	0	0	48	2	4.040	0.196
56399613	interior_score	0	0	0	4	46	4.920	0.271
56399613	exterior_score	0	0	0	47	3	4.060	0.237
56402163	interior_score	0	0	10	40	0	4.800	0.400
56402163	exterior_score	0	0	6	44	0	4.120	0.325
56389956	interior_score	0	0	0	0	50	5.000	0.000
56389956	exterior_score	0	0	0	44	6	4.880	0.337
56384305	interior_score	0	0	36	14	0	4.280	0.449
56384305	exterior_score	0	0	45	1	0	3.940	0.310
56391263	interior_score	0	0	0	50	0	5.000	0.000
56390649	interior_score	0	0	0	32	18	4.360	0.480
56390649	exterior_score	0	0	45	5	0	4.310	0.300
56388347	interior_score	0	0	0	50	0	5.000	0.000
56388347	exterior_score	0	0	42	7	0	4.120	0.382

Table 4.7: Composite Model Score Mean & Std

zpid	dimension	score_1	score_2	score_3	score_4	score_5	mean_score	std_score
56403089	interior_score	0	0	0	42	8	4.160	0.367
56403089	exterior_score	0	0	1	46	3	4.040	0.280
56400471	interior_score	0	0	0	47	3	4.060	0.237
56400471	exterior_score	0	0	4	46	0	3.920	0.271
56399613	interior_score	0	0	0	49	1	4.020	0.140
56399613	exterior_score	0	0	0	50	0	4.000	0.000
56402163	interior_score	0	0	0	49	1	4.020	0.140
56402163	exterior_score	0	0	0	50	0	4.000	0.000
56389956	interior_score	0	0	2	48	0	3.960	0.196
56389956	exterior_score	0	0	2	48	0	3.960	0.196
56384305	interior_score	0	1	48	1	0	4.000	0.200
56384305	exterior_score	0	6	44	0	0	3.880	0.325
56391263	interior_score	0	0	5	45	0	4.900	0.300
56391263	exterior_score	0	0	0	50	0	4.980	0.140
56390649	interior_score	0	0	0	50	0	4.000	0.000
56390649	exterior_score	0	0	0	50	0	4.000	0.000
56388347	interior_score	0	0	0	47	3	4.060	0.237
56388347	exterior_score	0	0	2	48	0	3.960	0.196

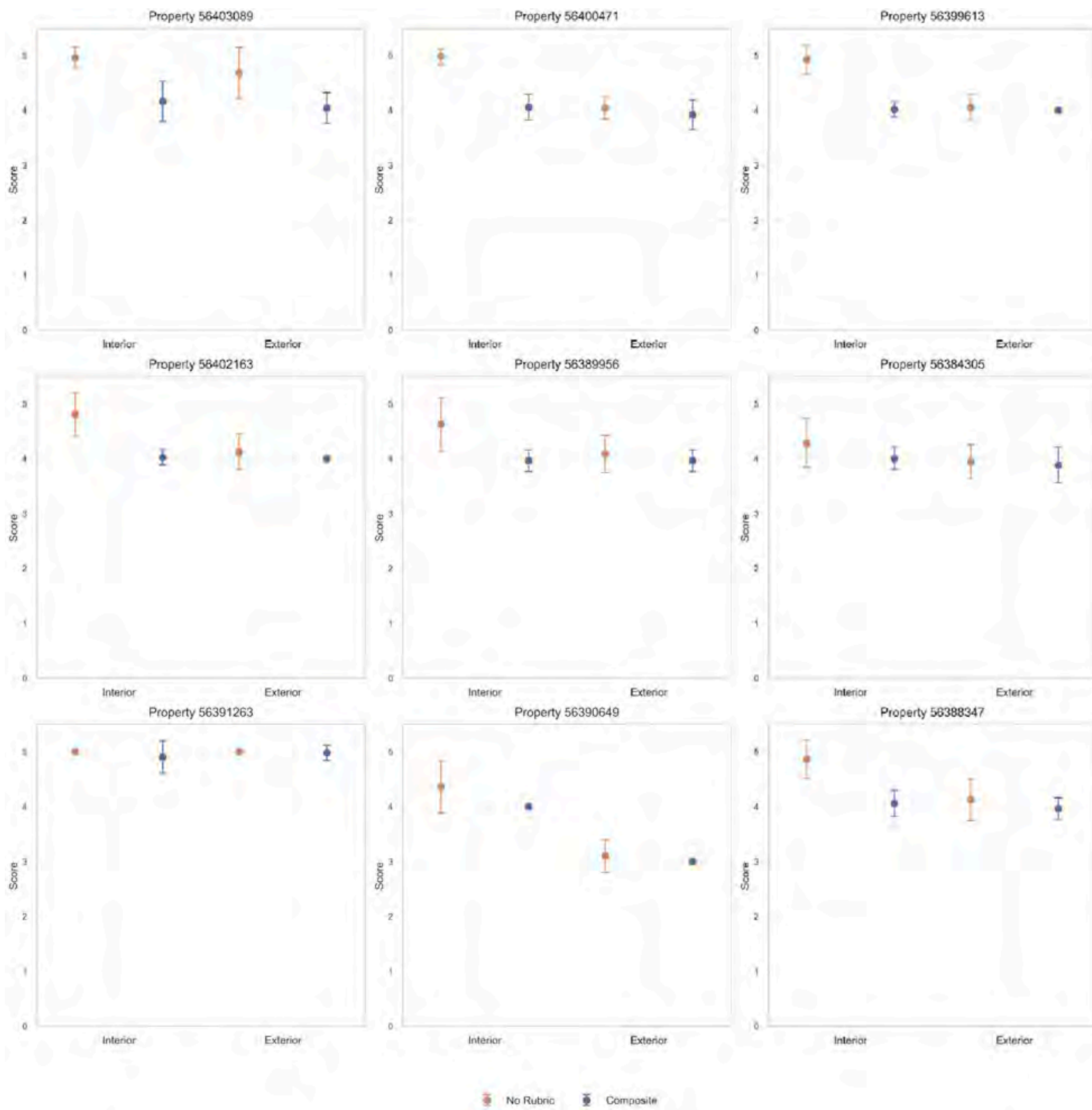


Figure 4.4: No_Rubric Score Mean & Std vs. Composite Score Mean & Std

In Fig. 4.4, the x-axis represents the interior score and exterior score for both No_Rubric method in orange and Composite method in blue, and the y-axis shows 1-5 score. Each circle marks the mean score value, while the lines indicate the range of standard deviation, providing a visual representation of the score variability.

The analysis shows that the composite method produces more stable and consistent AI-generated scores compared to the No_Rubric method, as evidenced by consistently lower standard deviations (std) across all 9 properties. For interior scores, the composite method has an average std of 0.19, compared to 0.33 in the No_Rubric method, while for exterior scores, the composite method achieves an average std of 0.16, significantly lower than the 0.30 in the No_Rubric method. Additionally, the range of std values in the composite method is narrow, from 0.00 to 0.37 for interior scores and 0.00 to 0.32 for exterior scores, further demonstrating that the composite method provides a highly stable scoring system with minimal variability across multiple evaluations.

The results highlight the importance of a well-defined rubric in achieving better performance. Without a rubric, as in the No-Rubric model, the scoring process becomes subjective and inconsistent, leading to higher variability in predictions. This is analogous to asking a person to score 300 properties without clear guidelines—while general patterns might be captured, the scores would likely fluctuate significantly due to individual biases or fatigue. Furthermore, it becomes very challenging to track the scoring process along the way, as the scorer might lose sight of the standards applied to earlier properties. As a result, each property is judged relative to the previous one, with scores capturing differences from the immediately preceding property rather than adhering to a consistent standard. Interestingly, the No-Rubric model tends to assign higher scores than the Composite model, which might reflect a leniency bias when lacking structured criteria. This could stem from the AI system's attempt to interpret and reward visual elements broadly, without penalizing certain deficiencies rigorously.

Chapter 5

Discussion

5.1 Investor Confidence and Development Guidance

A more accurate valuation method can enhance transparency by providing reliable, data-driven property assessments that reduce informational discrepancies across the market. In examining the economic impact of transparency on real estate markets, the JLL and LaSalle underscores the role that transparency plays in attracting investment and enhancing liquidity [3]. The report reveals that highly transparent real estate markets, including those in the U.S., U.K., Canada, and Australia, have collectively drawn over 1.2 trillion in direct commercial real estate investment over the last two years, representing more than 80 percent of the global total. This concentration of investment highlights the financial benefits of transparency, as markets with clear pricing structures and accessible data provide a reduced-risk environment that build investor confidence and accelerates transaction flows. Additionally, these transparent markets are well-positioned to lead liquidity recoveries in real estate cycles, particularly during periods of economic volatility. These findings suggest that increasing transparency can contribute to a more stable and attractive real estate market, creating long-term value for investors and supporting broader economic resilience.

Furthermore, a more transparent market can creates greater market diversity by enabling participation from a broader range of stakeholders, including small agencies, individual buyers, and emerging investors. By providing clear, accessible data and reducing informational asymmetries, transparency lowers entry barriers and encourages a more diverse set of players to engage in the real estate market. This diversity not only enhances competition but also drives innovation, as a wider array of perspectives and needs influences market dynamics, creating a more vibrant and inclusive real estate ecosystem.

Figure 5.1: Institutional Investment vs Market Data Availability for Emerging Sectors. Source: JLL Global Real Estate Transparency Index 2024.

The included diagram titled "Institutional investment vs market data availability for emerging sectors" further illustrates the strong link between transparency and institutional investment. Emerging real estate sectors such as student housing, data centers, and senior housing are shown to have high levels of institutional involvement in highly transparent markets, underscoring the significant role of reliable data availability in attracting institutional investors. In transparent markets, investors have access to clear information on property values, operational costs, and market trends, which helps them better assess potential returns and risks. This data-driven approach aligns with institutional investors' need for consistent, predictable returns, as they often manage large portfolios on behalf of pension funds, endowments, and other entities that prioritize long-term stability.

Implementing an accurate buyer-centric valuation model can significantly enhance real estate development decisions by aligning projects with actual buyer preferences, thereby increasing market appeal and reducing the risk of unsold inventory. By integrating data on buyer desires—such as preferred home features, layouts, and amenities—developers can tailor properties to meet market demand more precisely. For instance, National Association of Home Builders reveals that 85% of buyers prefer an open arrangement between the kitchen and dining room, and 79% between the kitchen and family room. Incorporating such insights into development plans ensures that new properties resonate with prospective buyers, leading to higher satisfaction and quicker sales [12]. The study from Crosby emphasizes the importance of understanding market demand and buyer preferences in achieving successful real estate investments [13]. It highlights that aligning property valuations with the actual desires of buyers not only enhances the appeal of new developments but also minimizes the risk of unsold inventory by catering directly to what buyers seek.

5.2 Limitation

One of the primary limitations of the current valuation method lies in its dependency on the ChatGPT-4o model for scoring images based on visual property attributes. ChatGPT-4o, a closed-source model, presents several challenges due to the lack of transparency in its training process. Specifically, because OpenAI does not disclose detailed information on the model's training data or architecture, it is unclear how changes in the model architecture or training set influence its scoring capabilities over time. Consequently, each time OpenAI releases an update model, there is an inherent risk that the scoring criteria may shift, affecting the consistency and comparability of the generated scores. From the study by Smith, it is evident that certain models require periodic re-evaluation or retraining to maintain accuracy [14]. However, each update can alter the model's behavior, potentially impacting the reliability of valuations. In fields like real estate, where consistent and dependable valuations are crucial, the study highlights the need for transparent and adaptable models to effectively handle these updates.

Another limitation of the current valuation method is its narrow scope in both the time and space dimensions. The analysis is restricted to data from 2023, which may not account for variations in market dynamics across different years or economic cycles. As real estate markets are influenced by factors such as economic growth, interest rates, and policy changes, extending the temporal scope is essential to assess the model's robustness over time. Additionally, the study focuses exclusively on properties in Arlington and Woburn, Massachusetts, limiting its applicability to these specific markets. Real estate markets vary significantly across regions due to differences in demographics, local regulations, and market conditions. To test the generalizability and broader applicability of the method, future work should include data from diverse locations and multiple years, enabling a more comprehensive evaluation of the model's performance across varied contexts.

5.3 Future to Explore

A promising direction for future exploration is incorporating apartment views into the valuation model. Views play a significant role in property desirability and, subsequently, rental and sale values, particularly in urban settings. In the study by Chang and Lee, the researchers examined the influence of scenic views on property values, particularly in urban apartment settings [15]. They found that apartment units with desirable views, such as cityscapes, waterfronts, or green spaces, consistently commanded higher sale and rental prices compared to those without such views. This "view premium" was especially pronounced in densely populated urban areas, where limited access to scenic views made them highly sought after and increased property appeal. Incorporating a "view quality" score into the model for apartments would enhance its applicability in rental markets, where views are often a significant determinant of rental pricing. By developing a scoring mechanism to quantify the quality of apartment views, the model could be expanded to capture a critical variable in the valuation of rental properties.

Another exciting avenue for exploration is evaluating the impact of renovations on property values. Understanding how pre- and post-renovation conditions affect market value would be particularly beneficial for investors, as it would help them assess the financial return of prospective renovations. National Association of Realtors provides insights into how specific renovations impact property values [16]. The report highlights that certain home improvements, particularly in key areas like kitchens and bathrooms, yield substantial returns on investment, with increased property value often exceeding the renovation costs. For example, kitchen remodels and bathroom upgrades are shown to enhance buyer appeal and contribute significantly to property valuation, although the value added may vary by region and market conditions. By scoring images of properties before and after renovations, the model could estimate the value added through renovation efforts, aiding investors in making informed decisions regarding renovation viability.

Chapter 6

Conclusion

This thesis aims to enhance the traditional hedonic model by incorporating visual property photo scores generated using ChatGPT-4o. The refined model is designed to better capture buyer preferences, resulting in valuations that align more closely with actual sold prices. Three different models were tested: the No_Rubric Model, the Composite Model, and the Verbose Model, with comparisons drawn against the base hedonic model. These variations were analyzed to determine whether the provided scoring rubric impacts performance and whether a more nuanced scoring system improves the model's predictive accuracy.

The dataset was divided into training and test sets, with performance evaluated using metrics such as test R^2, adjusted R^2, MAE, RMSE, and MAPE. Results indicate that the Composite Model performed the best, achieving the highest test R^2 of 86.4%, the lowest MAPE of 10.6%, and the lowest standard deviation of 0.088. It also outperformed the other models in terms of adjusted R^2, MAE, and RMSE on test data. While the No_Rubric and Verbose Models performed well on training data, their performance dropped significantly on test data, with higher standard deviations. This suggests that without a scoring rubric, ChatGPT's scores are less consistent, leading to greater variability. Conversely, the Verbose Model's use of overly detailed scoring criteria may lead to overfitting, causing unstable performance that depends heavily on the training-test dataset split.

The significance of this work lies in its potential to improve property valuation methods by integrating AI-driven qualitative assessments, offering a more buyer-centered perspective. This approach not only enhances pricing accuracy but also provides a framework for using AI to quantify subjective property attributes, paving the way for broader applications in real estate analytics.

Appendix A

Prompt

No_Rubric Prompt:

```
You are an expert in architecture and real estate photography. Your task is to evaluate the following property interior photos and give an interior score from 0 to 5, if there's no interior photo, give 0 for interior score; evaluate the following property exterior photos and give an exterior score from 0 to 5, if there's no exterior photo, give 0 for exterior score.

Reply in JSON format:
{
    "interior_score": <integer from 0 to 5>,
    "exterior_score": <integer from 0 to 5>,
    "explanation": "<explanation for the score in one paragraph, do not use line return>"
}
```

Listing A.1: No_Rubric Prompt

Composite Prompt:

```
You are an expert in architecture and real estate. Your task is to
    evaluate the property photo based on aesthetics appeal,
    cohesiveness, and condition using the following rubric:

Scoring Rubric for Interior Photos

    Score: 0 (Missing)
    If there are no interior photos, use 0. Do not use null or NaN.

    Score: 1 (Poor)
    Cohesiveness: The interior features a mix of design styles,
        colors, or materials that clash, creating a disjointed and
        chaotic appearance.
    Condition: The interior shows significant signs of neglect or
        damage, such as cracked walls, peeling paint, broken
        fixtures, or damaged flooring. Major repairs or renovations
        are needed, and the overall impression is of disrepair.
    Aesthetic Appeal: The space is visually unappealing, with
        little attention to design or decoration, resulting in a
        lackluster interior.

    Score: 2 (Fair)
    Cohesiveness: Some mismatched styles or elements exist, but the
         overall interior still holds together, though imperfectly.
    Condition: The interior is functional but shows wear and tear,
        with minor damage like chipped paint or outdated fixtures.
        Cleanliness is subpar, and the space needs updates for
        better presentation.
    Aesthetic Appeal: Some elements are visually pleasing, but the
        overall aesthetic lacks balance and refinement.

    Score: 3 (Average)
    Cohesiveness: The interior follows a single design style but
        lacks strong thematic unity or creativity.
    Condition: The interior is in decent condition, with minor
        cosmetic issues like slight wear but overall clean and well-
        maintained.
```

```
    Aesthetic Appeal: The interior has some visual appeal but lacks standout features, resulting in a pleasant but unremarkable space.

    Score: 4 (Good)
    Cohesiveness: The design shows a clear and consistent style with well-coordinated materials, colors, and d cor elements.
    Condition: The interior is well-maintained and clean, with minimal visible wear or imperfections.
    Aesthetic Appeal: The space is visually appealing, with thoughtful design choices that create an attractive environment.

    Score: 5 (Excellent)
    Cohesiveness: The interior is perfectly cohesive, with every design element working together harmoniously.
    Condition: The interior is in immaculate condition, appearing new or recently renovated with high-quality finishes.
    Aesthetic Appeal: The interior is visually stunning, with exceptional attention to detail, creativity, and elegance.

Scoring Rubric for Exterior Photos

    Score: 0 (Missing)
    If there are no exterior photos, use 0. Do not use null or NaN.

    Score: 1 (Poor)
    Cohesiveness: Disjointed design with clashing architectural styles or materials.
    Condition: Significant visible damage or neglect, such as broken windows or overgrown vegetation.
    Aesthetic Appeal: The property photo is visually unappealing, showing neglect or lack of effort in presentation.

    Score: 2 (Fair)
    Cohesiveness: Some stylistic inconsistencies exist, but the design feels somewhat connected.
```

```
    Condition: Noticeable wear and tear, with minor damage like chipped paint or outdated landscaping.
    Aesthetic Appeal: The photo shows effort but lacks visual appeal, with poor composition or lighting.

    Score: 3 (Average)
    Cohesiveness: Consistent architectural style but lacking strong integration or thematic unity.
    Condition: Well-maintained overall, with minor cosmetic issues.
    Aesthetic Appeal: The photo is decent, showing the property in a positive light but without standout features.

    Score: 4 (Good)
    Cohesiveness: Clear and consistent architectural style with well-coordinated elements.
    Condition: Well-maintained property with thoughtful landscaping and minimal visible wear.
    Aesthetic Appeal: The photo is well-composed, capturing key features that enhance the propertys appeal.

    Score: 5 (Excellent)
    Cohesiveness: Perfectly cohesive design, with every element working together harmoniously.
    Condition: Pristine condition with no visible flaws. High-quality landscaping adds significant curb appeal.
    Aesthetic Appeal: The photo is visually stunning, highlighting the propertys best features with expert composition.

Reply in JSON format:
{
    "interior_score": <integer from 0 to 5>,
    "exterior_score": <integer from 0 to 5>,
    "explanation": "<explanation for the score in one paragraph, do not use line return>"
}
```

Listing A.2: Composite Prompt

Verbose Prompt:

```
You are an expert in architecture and real estate. Your task is to evaluate the following property photo based on aesthetic appeal, cohesiveness, and condition using the following rubric:

Interior Cohesiveness:
    Score: 0 (Missing) - If there is no interior photos, use 0, do not use null or NaN
    Score: 1 (Poor) - The interior features a mix of design styles, colors, or materials that clash, creating a disjointed and chaotic appearance.
    Score: 2 (Fair) - There are some mismatched styles or elements, but the overall interior still holds together, though imperfectly.
    Score: 3 (Average) - The interior follows a single design style, but lacks strong thematic unity or creativity. The style is cohesive but unremarkable.
    Score: 4 (Good) - The interior design shows a clear and consistent style with well-coordinated materials, colors, and d cor elements that complement each other.
    Score: 5 (Excellent) - The interior is perfectly cohesive, with every design element colors, materials, and styles working together seamlessly to create a unified, harmonious space.

Interior Condition:
    Score: 0 (Missing) - If there is no interior photos, use 0, do not use null or NaN
    Score: 1 (Poor) - The interior shows significant signs of neglect or damage, such as cracked walls, peeling paint, broken fixtures, or damaged flooring. The space looks cluttered or dirty, with little to no care taken to maintain the condition. Major repairs or renovations are clearly needed, and the overall impression is one of disrepair and poor upkeep.
    Score: 2 (Fair) - The interior is functional but shows noticeable wear and tear. There may be minor damage, such as chipped paint, outdated fixtures, or stained flooring.
```

```
    While generally usable, the space could benefit from cosmetic improvements. Cleanliness may be subpar, and the overall condition suggests the need for some repairs or updates to make the space more presentable and comfortable.
    Score: 3 (Average) - The interior is in decent condition, with only minor cosmetic issues. There may be some slight wear, but the fixtures, walls, and floors are all in good working order. The space appears clean and well-maintained, though not necessarily modern or freshly updated. No major repairs are needed, and the overall impression is of a lived-in, functional, and well-kept environment.
    Score: 4 (Good) - The interior is well-maintained and in great condition. There are few, if any, visible signs of wear. Fixtures, flooring, and walls look clean and in good working order. Any imperfections are minor and easily overlooked. The space appears fresh and well cared for, with only minor updates needed to make it look pristine.
    Score: 5 (Excellent) - The interior is in immaculate condition, with no visible flaws. Everything appears new or recently renovated, including high-quality fixtures, flooring, and finishes. The space is spotless, well-organized, and visually appealing, exuding a sense of luxury or care. No repairs or updates are necessary, and the overall impression is of a perfectly maintained and aesthetically pleasing interior.

Interior Aesthetic Appeal:
    Score: 0 (Missing) - If there is no interior photos, use 0, do not use null or NaN
    Score: 1 (Poor) - The space is visually unappealing, with little attention to design or decoration, resulting in a lackluster and unattractive interior.
    Score: 2 (Fair) - While certain elements may be visually pleasing, the overall aesthetic lacks balance and refinement, and the space feels underwhelming.
    Score: 3 (Average) - The interior has some visual appeal, but lacks striking features or a strong design statement. It is pleasant but not particularly memorable.
```

```
    Score: 4 (Good) - The space is visually appealing, with
       thoughtful design choices that create an attractive and
       comfortable environment.
    Score: 5 (Excellent) - The interior is visually stunning, with
       exceptional attention to detail, creativity, and elegance.
       Every design choice contributes to a beautiful and
       sophisticated space.

Exterior Cohesiveness:
    Score: 0 (Missing) - If there is no exterior photos, use 0, do
       not use null or NaN
    Score: 1 (Poor) - The design is disjointed, with clashing
       architectural styles or materials that create a chaotic,
       incoherent appearance.
    Score: 2 (Fair) - Some stylistic inconsistencies exist, but the
        overall design still feels somewhat connected, though
       imperfect.
    Score: 3 (Average) - The architectural style is consistent, but
        lacks strong integration or thematic unity. The design is
       cohesive but unremarkable.
    Score: 4 (Good) - The design shows a clear and consistent
       architectural style with well-coordinated materials and
       elements that complement each other.
    Score: 5 (Excellent) - The design is perfectly cohesive, with
       every architectural element working together seamlessly to
       form a unified, harmonious structure.

... (similar for Exterior Condition and Exterior Aesthetic Appeal)

Reply in JSON format:
{
    "interior_cohesiveness_score": <integer from 0 to 5>,
    "interior_condition_score": <integer from 0 to 5>,
    "interior_aesthetic_appeal_score": <integer from 0 to 5>,
    "exterior_cohesiveness_score": <integer from 0 to 5>,
    "exterior_condition_score": <integer from 0 to 5>,
    "exterior_aesthetic_appeal_score": <integer from 0 to 5>,
    "explanation": "<explanation for the score in one paragraph, do
```

```
        not use line return>"
}
```

Listing A.3: Verbose Prompt

References

[1] S. B. Kask and S. A. Maani. “Uncertainty, Information, and Hedonic Pricing”. In: *Land Economics* 68.2 (1992), pp. 170–184.

[2] A. C. Goodman. “Hedonic Prices, Price Indices and Housing Markets”. In: *Journal of Urban Economics* 5.4 (1978), pp. 471–484.

[3] JLL & LaSalle. *Global Real Estate Transparency Index, 2024: Transparency in a Time of Transition*. Tech. rep. 2024.

[4] S. Rosen. “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition”. In: *The Journal of Political Economy* 82.1 (1974), pp. 34–55.

[5] R. Halvorsen and H. O. Pollakowski. “Choice of Functional Form for Hedonic Price Equations”. In: *Journal of Urban Economics* 10 (1981), pp. 37–49.

[6] O. Poursaeed, T. Matera, and S. Belongie. “Vision-Based Real Estate Price Estimation”. In: *Machine Vision and Applications* 29.4 (2018), pp. 667–676. DOI: 10.1007/s00138-018-0922-2.

[7] Q. You, R. Pang, L. Cao, and J. Luo. “Image-Based Appraisal of Real Estate Properties”. In: *IEEE Transactions on Multimedia* 19.12 (2017), pp. 2751–2759. DOI: 10.1109/TMM.2017.2710804.

[8] A. Dosovitskiy et al. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”. In: *arXiv preprint arXiv:2010.11929* (2020).

[9] M. Yazdani and M. Raissi. “Real Estate Property Valuation Using Self-Supervised Vision Transformers”. In: *arXiv preprint arXiv:2302.00117* (2023).

[10] J. D. Kintzel. “Price Prediction and Deep Learning in Real Estate Valuation Models”. MA thesis. Harvard University, 2019.

[11] Zillow. *Zillow Real Estate Website*. Accessed: 2024-11-21. 2024. URL: https://www.zillow.com.

[12] National Association of Home Builders. *What Home Buyers Really Want, 2024 Edition*. BuilderBooks, 2024.

[13] N. Crosby. “Valuation Accuracy, Variation and Bias in the Context of Standards and Expectations”. In: *Journal of Property Investment & Finance* 18.2 (2000), pp. 130–161. DOI: 10.1108/14635780010324320.

[14] J. Smith, L. Brown, and M. Davis. “Evaluating Model Transparency and Adaptability for Reliable AI-Driven Real Estate Valuations”. In: *Journal of Machine Learning in Real Estate* 12.3 (2023), pp. 45–67.

[15] T. Chang and K. Lee. “The Impact of Scenic Views on Property Valuation: A Study of Urban Apartments”. In: *Journal of Real Estate Research* 48.2 (2022), pp. 134–156.

[16] National Association of Realtors. *Home Buyer and Seller Generational Trends Report*. Tech. rep. National Association of Realtors Research Group, 2024.