

Article

Benefit Evaluation of Urban Commercial Land and XGBoost-SHAP Based Influence Analysis: A Case Study of Chengdu, China

Yuan Jiang ^{1,2}, Peng Tang ², Shijie Sun ^{1,*} , Ying Liu ¹ , Xiaorong Zhang ², Benying Xu ² and Gaomeiyuan Zheng ³

¹ School of Architecture, Southeast University, Nanjing 210096, China; 230249515@seu.edu.cn (Y.J.); yingliu@seu.edu.cn (Y.L.)

² Department of Digital Planning, Chengdu Institute of Planning & Design, Chengdu 610041, China; swanytang@hotmail.com (P.T.); zhangxiaorong16@mails.ucas.edu.cn (X.Z.); xubenying11@163.com (B.X.)

³ SHUDAO Investment Group Co., Ltd., Chengdu 610095, China; zg_sd1993@163.com

* Correspondence: sun_shijie@seu.edu.cn

Abstract

Accurately assessing the multifaceted performance of urban commercial land is essential for steering consumption upgrading and refining urban spatial planning. However, existing scholarship continues to treat commercial land through a predominantly economic lens; non-economic indicators have rarely been compared with price-based metrics within a unified analytical framework. To bridge this gap, this study constructs a multidimensional benefit evaluation framework encompassing four indicators, namely, land values, shop rent, online reviews, and customer ratings, across 10,766 commercial land parcels in Chengdu, China, and applies XGBoost-SHAP methods to examine their predictive associations and spatial patterns. This study reveals the differences and commonalities among the four benefit indicators in terms of statistical characteristics, spatial distribution, and matching patterns, and further examines the ranking of their key predictors, as well as the nonlinear effects, threshold effects, and interaction effects with key factors. These findings corroborate the distinctiveness of the four benefit variables and the significance of comprehensively evaluating commercial land-use benefits. This study provides novel perspectives and empirical evidence for assessing commercial land-use benefits and their predictive mechanisms, and offers actionable guidance for formulating targeted consumption development, spatial layout, and business format optimization strategies, thereby contributing to the sustainable vitality of urban economic spaces.

Keywords: commercial land; benefit evaluation; influence analysis; XGBoost-SHAP; Chengdu



Academic Editor: Bindong Sun

Received: 6 May 2026

Revised: 23 June 2026

Accepted: 25 June 2026

Published: 27 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The assessment of urban land benefits has long served as a foundational analytical tool for informing land-use allocation decisions and enhancing spatial efficiency [1–3]. Unlike residential or industrial land, where benefits can be reasonably proxied by transaction prices, tax revenues, and other economic data [4–6], the performance of a commercial parcel reflects the combined outcomes of economic returns, utilization intensity, consumer experience, and online visibility in increasingly digitalized markets. However, the bulk of existing scholarship continues to treat commercial land through a predominantly economic lens, measuring success in terms of land values [1,2,7], shop rents [8], or the capitalization

effects of transit investments [9,10]. The non-economic dimensions of commercial land benefits, such as customer vitality [11,12], customer reviews [13,14] and others, have received comparatively scant systematic attention.

In recent years, the commercial real estate markets across global cities have confronted structural headwinds: rising vacancy rates [12] and softening rental yields [11,15]. According to reports, in 2024, the average vacancy rate of famous shopping centers in the central urban areas of major Chinese cities was 16.5%, and rental income levels decreased by an average of 21.5%, marking a meaningful departure from the near-full-occupancy norms of the preceding decade. However, during the same timeframe, owing to the ongoing diversification of business formats, the steady enhancement of business capabilities, and the continual refinement of commercial space quality, Chinese consumers have demonstrated heightened enthusiasm and preference for commercial spaces [16,17]. These evolving market dynamics raise a critical empirical question that conventional single-indicator evaluation frameworks cannot adequately address: how to systematically assess the multifaceted benefits of commercial land use and the extent to which economic and non-economic benefit indicators converge or diverge in their quantitative magnitudes, spatial distribution patterns, key influencing factors, and underlying mechanisms.

The paper addresses this question by using four benefit indicators of commercial land, namely, land value, shop rent, customer rating and online reviews, across 10,766 commercial land parcels in Chengdu, China. First, this article tests whether the four benefit indicators exhibit convergent or divergent geographic signatures. Second, the relative importance of the 24 independent variables across four dimensions (location conditions, surrounding conditions, travel conditions, and construction status) and the four benefit indicators is sequentially examined using SHAP-derived feature importance rankings. Third, this study characterizes whether threshold effects and marginal response patterns of key variables vary across benefit indicators, yielding dimension-specific policy implications. Methodologically, XGBoost (eXtreme Gradient Boosting) [18] with SHAP (SHapley Additive exPlanations) value decomposition [19] is employed to model each benefit indicator. Although spatial blocking cross-validation is the preferred evaluation strategy for spatial prediction tasks, the primary objective of this study is attribution, identifying which factors are predictively associated, rather than spatial prediction. Five-fold random cross-validation is therefore adopted to evaluate model fit, as random partitioning allows all geographic regions to contribute to both training and testing, yielding stable estimates of feature importance. All models are trained on a unified set of 24 independent variables, ensuring comparability of feature importance rankings across benefit types.

The remainder of this article proceeds as follows. Section 2 reviews the relevant literature on commercial land benefit evaluation. Section 3 shows the study area, data sources, and variable construction. Section 4 shows the methodological framework. Section 5 presents the results and discusses these findings. Section 6 concludes this study.

2. Literature Review

2.1. Benefit Evaluation: From Land Price to Multidimensional View

The intellectual foundations of commercial land benefit evaluation trace directly to Alonso's bid-rent theory [5], which established accessibility to the Central Business District (CBD) as the principal determinant of a parcel's capacity to generate commercial returns. For decades, this framework led scholarship to treat "commercial land benefit" as effectively synonymous with land transaction prices or rental levels, with empirical work relying predominantly on hedonic price models that regress observed prices or rents against structural and locational attributes [20,21]. This unidimensional focus had practical grounds, as price data were comparatively accessible.

Some developments have gradually expanded the scope of benefit evaluation beyond this price-centric view. First, a growing body of work has treated the performance of a commercial parcel as a multidimensional construct encompassing utilization intensity [10], visitor satisfaction, digital visibility in online platforms [22,23], and other factors. Liu et al. [22] explicitly argued that these dimensions should be treated as distinct rather than proxyable indicators, demonstrating that supply–demand relationships in commercial land markets cannot be captured by land values alone. Similarly, Gebru et al. [12] showed that vacancy, a direct indicator of utilization failure, was poorly predicted by land values in their analysis of vacant urban land in Ethiopia. Second, the proliferation of crowd-sourced platforms (e.g., Dianping and Meituan) and street-view imagery has made it feasible to construct parcel-level indicators for non-economic indicators that were previously unmeasurable at scale [24–26]. Recently, an increasing number of studies have examined some non-economic dimensions, such as customer satisfaction [13] and online visibility [27], rather than systematically comparing multiple benefit types within a unified analytical framework. However, limited research has examined whether economic and non-economic benefit indicators exhibit convergent or divergent geographic patterns, and whether their driving factors differ fundamentally in identity, direction, relative importance, threshold effects and interaction patterns.

2.2. *Influencing Factors: Using the Most Widely Cited Factors*

The driving factors of commercial land benefits and related phenomena, such as transit-oriented development (TOD), can be systematically categorized using the 3D (density, diversity, and design)/5D (adding distance to transit and destination accessibility) framework [28,29], which encompasses a comprehensive set of variables. To synthesize existing research and streamline the analysis, this paper classifies the most widely cited factors in the literature into four dominant dimensions: location conditions, surrounding conditions, travel conditions, and construction status. Notably, additional factors such as policy interventions [4,21] and institutional arrangements have been explored in prior studies.

Location conditions constitute the most extensively studied category, grounded in bid-rent theory and central place logic [30]. Distance to the CBD, proximity to designated commercial centers, and position within the urban hierarchy have consistently emerged among strong predictors of land values and rent levels [3]. More recent work has incorporated plan-designated hierarchies by the local government, such as the Urban Commercial District (UCD) and Economic Development Zone (EDZ), as policy-oriented location variables that capture agglomeration and political externalities beyond pure distance metrics [22,23].

Surrounding conditions usually encompass the socio-economic and land-use characteristics around the commercial parcel. Population density and income define the scale and purchasing power of the potential customer base [10,13,31]; LUM determines the stability and diversity of land use [29,32]; and public service amenities, such as schools, hospitals, and parks, generate spillover effects that shape shopping mall attractiveness. A growing body of work has also emphasized perceptual streetscape qualities [25,26,33], including the green view index (GVI), landscape diversity, and openness, as determinants of commercial vitality that operate through psychological and experiential pathways rather than purely market-driven ones [25,34].

Travel conditions reflect how easily consumers can reach a given commercial destination. Proximity to metro stations with TOD has been shown to produce sizable positive effects on both commercial values and viability [9,27,35]. Road density and bus-stop coverage offer complementary measures of auto-oriented and transit-oriented accessibility,

respectively [7,15]. Walkability metrics, including intersection density [26] and pedestrian-space proportion [32], have attracted growing attention as evidence accumulates that pedestrian-friendly environments support stronger retail performance [28].

Construction status refers to site-intrinsic characteristics such as parcel size [35], building age [27], floor area ratio (FAR) [26], star rating [36], and business composition [37]. These factors may produce combined effects with variables from other dimensions, for instance, a well-located parcel in an aging, low-rise building may underperform one in a moderately good location with modern, vertically developed facilities [7].

The variables reviewed above have been repeatedly applied and validated across multiple cities and regional contexts. This study therefore adopts substantially the same variable framework, with adjustments governed by data availability in Chengdu. One consideration that warrants empirical attention is the sensitivity of factor importance to the spatial scale at which surrounding-condition variables are computed [38]. Consequently, a multi-scale sensitivity analysis is conducted in the modeling phase to determine the appropriate buffer size of commercial land.

2.3. Analytical Methods: Interpretable Machine Learning

The dominant analytical paradigm for land-benefit drivers has long been the hedonic price model [39] and its spatial extensions [40], geographically weighted regression and spatial econometric specifications. These methods impose linearity and additive separability assumptions that sit uneasily with the nonlinear, interactive character of urban systems. Tree-based ensemble methods, particularly XGBoost [18], have emerged as a powerful alternative. As a regularized gradient-boosting framework, XGBoost handles missing values natively, is robust to multicollinearity, and automatically captures high-order interactions without manual specification [18]. It has been successfully applied to land value estimation, housing rent modeling, and transit capitalization analysis, consistently outperforming linear benchmarks [41].

The principal objection to tree ensemble methods is their opacity: an XGBoost model aggregates predictions across hundreds of trees whose internal logic resists direct inspection. Among post hoc interpretability tools, LIME (Local Interpretable Model-agnostic Explanations) [42] provides local surrogate explanations but lacks global consistency, and ALE (Accumulated Local Effects) [43] captures average marginal effects robust to correlated predictors but does not support instance-level attribution. SHAP [19] differs from both by decomposing each prediction into additive feature contributions grounded in cooperative game theory, satisfying local accuracy, missingness, and consistency axioms [19]. This dual capability, including global importance ranking and local dependence structure, is essential when the research objective involves comparing driver hierarchies across multiple outcome dimensions.

Notably, when the research objective is spatial prediction, ordinary random cross-validation applied to spatial data allows training and test folds to contain geographically proximate observations, potentially inflating performance estimates. However, the primary objective of this study is influence rather than spatial prediction. Under random cross-validation, all geographic regions contribute to both training and testing, yielding stable estimates of feature importance rankings. This study therefore deploys separate XGBoost models across four benefit dimensions under 5-fold random cross-validation. To further assess the sensitivity of the results to the CV (cross-validation) strategy, SHAP-based driver rankings are also compared under spatial blocking validation; the consistency of rankings is reported as a robustness check. SHAP outputs are interpreted strictly as predictive associations rather than causal effects.

3. Research Area and Variables

3.1. Study Object and Scope

This study focused on commercial land parcels whose primary functions were retail, entertainment, shopping, and other consumption-oriented activities. Parcels serving predominantly industrial or finance functions, including hotels, wholesale markets, financial institutions, and office buildings, were excluded from the analysis. Parcel identification and functional classification were based on the 2024 Chengdu Land Survey Database, which recorded the designated land-use type and primary function for each parcel. After filtering, 10,766 parcels met the inclusion criteria.

The study area covered the central area of Chengdu (Figure 1), comprising 14 administrative districts. Among these, Jinjiang, Chenghua, Wuhou, Jinniu, and Qingyang formed the earliest-established and comprehensive functionalization areas, called legacy urban districts. Gaoxin and Tian Fuxinqu were designated in recent years as priority zones for integrated industry–city development. Qingbaijiang, Longquanyi, Shuangliu, and Pidu were manufacturing-intensive districts, each with a pronounced sectoral specialization in logistics, automotive industry, aviation, and electronics, respectively. Wenjiang, Xindu and Xinjin functioned primarily as high-quality residential districts.

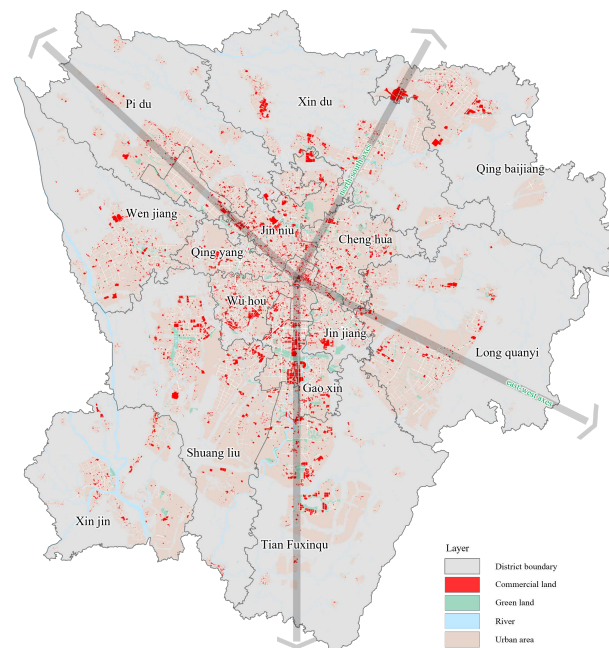


Figure 1. The study object and area.

Although these districts differed in economic structure, the tertiary sector accounted for over 50% of GDP in every district in 2024—exceeding 53% in Longquanyi, the least service-intensive—indicating that commercial and consumption-oriented activities constituted the dominant economic driver throughout the study area. All districts were therefore retained to capture the full heterogeneity of Chengdu’s commercial land use. Additionally, land, industrial, and fiscal resources within central Chengdu were predominantly concentrated along the city’s north–south and east–west axes.

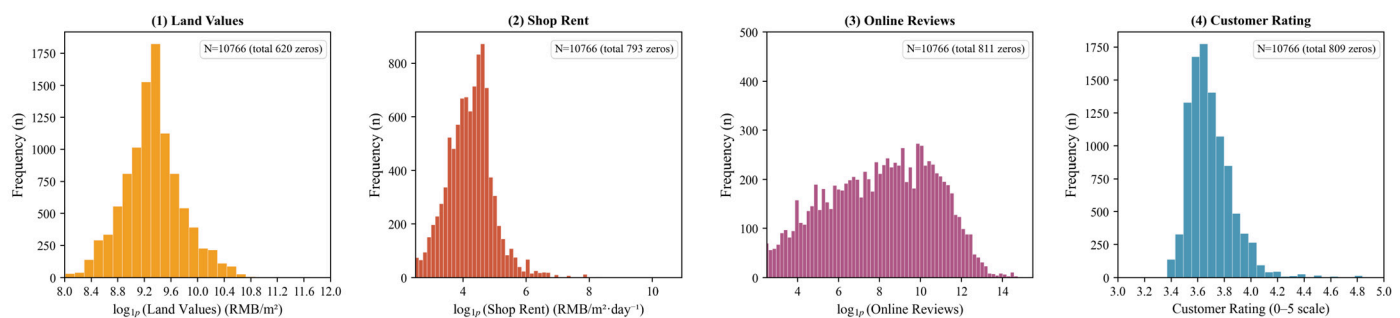
3.2. Dependent Variables

Two economic indicators, land values and shop rent, and two non-economic indicators, customer rating and online reviews, were used as dependent variables to assess the multifaceted benefits of commercial land (Table 1).

Table 1. Summary of dependent variables (unstandardized).

Variable	Description	Mean	Std.
Land values	Government-assessed benchmark price of commercial land (10^4 RMB/m ²)	1.17	0.661
Shop rent	Average of daily shop rents of commercial land (RMB/m ²)	92.01	375.91
Customer rating	Average of customer ratings (0–5) of commercial land	3.42	0.987
Online reviews	Monthly online reviews of commercial land (10^4)	3.16	11.16

Land values were obtained from the 2024 Chengdu Land Survey database as government-assessed benchmark land prices (10^4 RMB/m²) for commercial-use parcels. Among the 10,766 parcels, 620 (5.76%) recorded zero values, corresponding to parcels not yet priced in the official survey; these were retained to preserve full spatial coverage, and the variable was log-transformed as $\ln(1 + x)$ to accommodate the right-skewed distribution in Figure 2(1).

**Figure 2.** Distribution of 4 commercial land benefit variables.

Shop rent was sourced from the 2024 Chengdu Land Survey database as the listed daily rent (RMB/m²/day), aggregated to the parcel level by averaging all establishment-level values within each parcel. A total of 793 parcels (7.37%) had no rent listings and were assigned zero values; these were likewise retained to preserve full spatial coverage, and the variable was log-transformed as $\ln(1 + x)$ to accommodate the heavy right tail in Figure 2(2).

Online reviews were obtained from Dianping and Meituan as the monthly count of user-submitted reviews per parcel. A total of 811 parcels (7.53%) recorded zero reviews; these were retained to preserve full spatial coverage, and the variable was log-transformed as $\ln(1 + x)$ to accommodate the right-skewed distribution, in which a small subset of highly popular destinations accounted for the vast majority of the review volume in Figure 2(3).

Customer rating was obtained from Dianping and Meituan as the mean star rating (0.00–5.00) of establishments within each parcel; 809 parcels (7.51%) had no ratings and were assigned zero values. The mean was adopted because the narrow interquartile range (3.56–3.77) rendered the median insensitive to cross-parcel variation, and the robustness of this choice was verified by repeating key analyses with median-based ratings in Figure 2(4).

Notably, most customer online reviews and customer ratings are offline post-visit reviews and ratings, which effectively reflected the actual foot traffic and the consumer experience of commercial spaces, respectively.

3.3. Independent Variables

The independent variables were organized into four dimensions based on the 3D/5D built environment framework [28,29] in this paper, including location conditions, surrounding conditions, travel conditions, and construction status. A total of 24 indicators were identified (Table 2).

Table 2. Summary of independent variables (unstandardized).

Variable	Description	Mean	Std.
Location conditions			
Urban axis	0 = None, 1 = District-level secondary, 2 = District-level main, 3 = Municipal-level		
Urban center	0 = None, 1 = Area center, 2 = Regional center, 3 = Urban center		
UCD	0 = None, 1 = Grade III, 2 = Grade II, 3 = Grade I		
EDZ	0 = None, 1 = District-level, 2 = Municipal-level		
Surrounding conditions			
Population density	Employed and resident density within a 500 m radius (10 ⁴ people/km ²)	1.06	1.14
Income	Average income of population within a 500 m radius (RMB)	3790.7	1195.9
Job-housing index	Ratio of employed population to resident population within a 500 m radius	0.874	2.93
Average floors	Average number of floors within a 500 m radius (floors)	16.12	9.58
Public service	Proportion of public service land area within a 500 m radius (%)	7.34	8.09
Green space	Proportion of green land and park area within a 500 m radius (%)	4.12	6.58
Distinctive space	Proportion of distinctive space area within a 500 m radius (%)	1.25	6.28
LUM	Degree of land-use mix within a 500 m radius	0.856	0.189
Travel conditions			
Road density	Ratio of road length to area within a 500 m radius (km/km ²)	0.29	0.15
Subway station	Number of subway stations within a 500 m radius	0.26	0.43
Bus stops	Number of public transit stops within a 500 m radius	27.19	26.21
GVI	Average green view index of streets within a 500 m radius	0.111	0.077
Openness	Average openness index of streets within a 500 m radius	0.207	0.129
Pedestrian space	Proportion of pedestrian space in streets within a 500 m radius	0.028	0.022
Landscape diversity	Average landscape diversity index of streets within a 500 m radius	1.215	0.685
Construction status			
Area	Parcel area of the commercial land (ha)	0.907	2.458
Building height	Average height of the building on the commercial land (m)	18.98	7.25
Years	Years since opening of commercial land (years)	20.28	12.21
Star rating	Star rating (0–3) of the shopping mall on the commercial land	1.27	1.79
Business mix	Degree of commodity mix on the commercial land	0.078	0.133

Location conditions captured strategic spatial planning elements defined in Chengdu's Comprehensive Urban Plan. Urban axes were classified into four hierarchical tiers (municipal-level, district-level main, district-level secondary, and none), reflecting differentiated policy-related priorities and intended intensities of development. Urban centers were identified from the plan as designated sub-centers within the polycentric urban structure. UCD and EDZ designations were included as institutional location variables that capture policy-driven agglomeration externalities beyond pure distance metrics. The four variables recorded whether a parcel falls within designated planning zones regardless of buffer radius.

Surrounding conditions comprised eight indicators describing the socio-economic and land-use context of each parcel's immediate neighborhood. Population density and income were calculated to characterize the consumer base and purchasing power. Job-housing index and average building floors were included as proxies for sustained consumption vitality. Public service amenities, green space, and distinctive space (including landmarks and cultural facilities) were quantified as the area radius of each land-use type within the buffer. LUM was computed using the Shannon entropy index to capture the diversity of land-use functions within the neighborhood. All of the eight indicators were calculated in a special catchment around the commercial parcels. Data pertaining to indicators, including population and income, were obtained from mobile phone signaling data and

the national economic census. Land-use-related data were sourced from the 2024 Chengdu Land Survey database.

Travel conditions captured the supply of mobility infrastructure within a buffer using three conventional indicators from the Chengdu Resident Travel Survey database, including road density, subway stations, and bus stops. Beyond basic accessibility, this paper also studied spatial quality by incorporating four streetscape-derived metrics using street-view data, including the green view index (GVI), openness, pedestrian space, and landscape diversity, all of which have been empirically linked to commercial performance in prior research [25,26,34]. And finally, construction status encompassed five site-intrinsic characteristics. Parcel area, building age, star rating, business mix, and building height were obtained from the 2024 Chengdu Land Survey database and POI (Point of Interest) database.

4. Methodology

4.1. Catchment Radius

A multi-scale sensitivity analysis was conducted at 300 m, 500 m, and 800 m, which are the most widely discussed catchment radii in urban built environment research, to determine the appropriate buffer scale. All buffers were computed as Euclidean (straight-line) distances from parcel centroids. Model fit improved monotonically from 300 m to 800 m for shop rent, customer rating and online reviews (Table 3). However, the marginal gain from 500 m to 800 m was negligible for land values. SHAP-based driver rankings were the most stable between 300 m and 500 m (Table 4). Moreover, 500 m corresponds to the standard 5–10 min walking catchment adopted in the built-environment literature and facilitates direct comparison with prior studies [7,23]. Based on these considerations, a 500 m buffer radius was selected for this analysis.

Table 3. R² comparison with 300 m, 500 m, and 800 m buffers.

Dependent Variable	300 m R ²	500 m R ²	800 m R ²	From 300 m to 500 m ΔR ²	From 500 m to 800 m ΔR ²
Land values	0.369	0.622	0.555	+0.253	−0.066
Shop rent	0.454	0.526	0.566	+0.073	+0.040
Customer rating	0.453	0.546	0.556	+0.093	+0.010
Online reviews	0.786	0.829	0.841	+0.043	+0.012

Table 4. SHAP stability with 300 m, 500 m, and 800 m buffers.

Scale Pair	300 m vs. 500 m τ	500 m vs. 800 m τ	300 m vs. 800 m τ
Land values	59.42%	42.03%	37.68%
Shop rent	49.28%	48.55%	34.06%
Customer rating	70.29%	55.80%	43.48%
Online reviews	68.12%	43.48%	30.43%
Average	61.78%	47.47%	36.41%

4.2. Spatial Autocorrelation Analysis

The Global Moran’s I was computed for each dependent variable before and after XGBoost modeling to evaluate the extent of residual spatial dependence (Table 5). All four raw outcomes exhibited significant positive spatial autocorrelation. After model fitting, residual spatial autocorrelation declined by 53.7–76.8%. Although XGBoost substantially reduced spatial dependence in all cases, residual spatial autocorrelation remained statistically significant across all four benefit indicators ($p < 0.001$), suggesting that while the models capture a large portion of spatial structure, some unexplained geographic cluster-

ing persists in the residuals. The reduction rates ranged from 53.7% (shop rent) to 76.8% (customer rating). As a supplementary robustness check, a spatial blocking cross-validation (2×3 grid) was also conducted (Table 6). The lower CV-R² values obtained under spatial blocking were to be expected: removing entire geographic blocks during training deprives the model of spatial context and is inconsistent with the attribution-oriented goal of this study. Because SHAP-based driver identification requires models trained on the full range of spatial conditions, random 5-fold CV was retained as the primary evaluation strategy. To further verify that the key predictors identified under random CV remain robust when spatial autocorrelation is controlled, SHAP-derived feature importance rankings between the full-sample (random CV) and spatial CV models were compared. Kendall's τ rank correlation coefficients ranged from 0.935 (customer rating and shop rent) to 0.971 (online reviews), with top-five predictor overlap of 80–100% across all four indicators. These high concordance rates indicate that although spatial blocking substantially degrades out-of-sample predictive performance, the relative importance of key predictors remains stable, supporting the reliability of SHAP-based attributions reported in this study.

Table 5. Moran's I results of 4 benefit variables.

Dependent Variable	Raw <i>I</i>	Raw <i>p</i>	Residual <i>I</i>	Residual <i>p</i>	Significant?
Land values	0.825	<0.001	0.367	<0.001	Yes
Shop rent	0.656	<0.001	0.304	<0.001	Yes
Customer rating	0.544	<0.001	0.126	<0.001	Yes
Online reviews	0.841	<0.001	0.304	<0.001	Yes

Table 6. Random CV vs. spatial CV comparison.

Dependent Variable	Random CV R ²	Spatial CV R ²	Change	Kendall τ	Top-5 Overlap
Land values	0.622	0.050	−0.572	0.957	100%
Shop rent	0.526	0.088	−0.438	0.935	80%
Customer rating	0.546	0.194	−0.352	0.935	100%
Online reviews	0.829	0.652	−0.177	0.971	100%

4.3. XGBoost-SHAP Model

Four separate XGBoost regressors were trained, one per benefit indicator. Fixed hyperparameters were adopted, following the default recommended settings in the XGBoost documentation [18]. Specifically, the following hyperparameters were used: *n_estimators* = 200, *max_depth* = 6, *learning_rate* = 0.05, *subsample* = 0.8, *colsample_bytree* = 0.8, and *random_state* = 42. Feature importance was ranked by mean absolute SHAP value. Nonlinear relationships and threshold effects were examined through SHAP dependence plots, which trace each feature's marginal contribution against its observed values. The SHAP value [19] is computed as

$$\varphi_{ij} = \sum_{S \subseteq \{x_{i1}, \dots, x_{ip}\} \atop \{x_{ij}\}} \left(\frac{|S|!(p - |S| - 1)!}{p!} \right) (f_{x_i}(SU\{x_{ij}\}) - f_{x_i}(S))$$

where φ_{ij} represents the contribution of the *j*th feature of dependent variable *i*, x_i represents the feature values' vector of the instance to be explained of dependent variable *i*, *p* is the number of corresponding independent variables, and $f_{x_i}(S)$ is the prediction of feature

values in subset S which are marginalized over features excluded from S [44]. Then, the SHAP value [45] can be defined by an additive feature attribution method as

$$F(y_j) = \varphi_{i0} + \sum_{j=1}^M \varphi_{ij} y_j$$

where F is the interpretation model, M represents the number of features, y_j indicates the presence of features for each dependent variable, φ_{ij} is the coefficient in the explanation model, and φ_{i0} is a constant. The overall framework of the paper is shown in Figure 3.

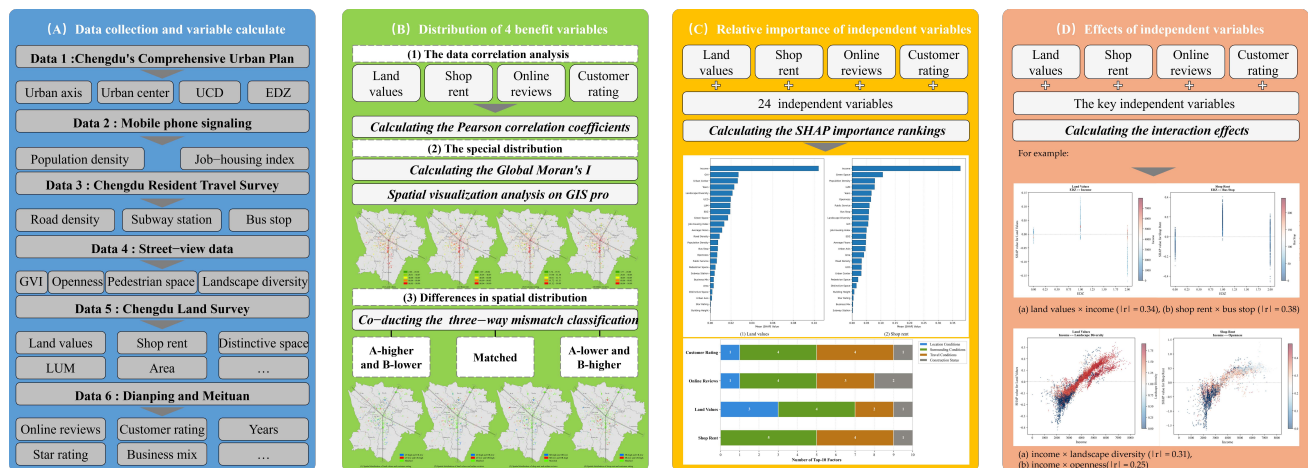


Figure 3. Overall framework of the paper.

5. Results and Discussion

5.1. Distribution of Four Benefit Variables

5.1.1. The Data Correlation Analysis

To characterize the relationships among the four benefit indicators, Pearson correlation coefficients were computed across all 10,766 commercial parcels in Figure 4.

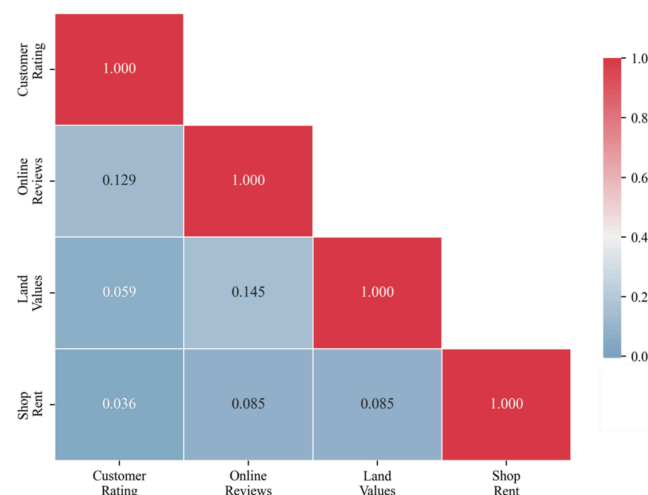


Figure 4. The Pearson correlation analysis.

All six Pearson pairwise correlations were low in magnitude ($|r| \leq 0.145$), with the highest coefficient observed between online reviews and land values ($r = 0.145$). Customer rating exhibited particularly weak linear associations with the two economic indicators, yielding correlations of only $r = 0.059$ with land values and $r = 0.036$ with shop rent. The correlation between the two economic indicators was similarly modest (land values–shop

rent: $r = 0.085$), as was the correlation between the two consumer-oriented indicators (customer rating–online reviews: $r = 0.129$). The uniformly low Pearson correlations indicate that the four benefit indicators capture substantially independent dimensions of commercial land performance, with linear overlap accounting for at most 2.1% of shared variance between any pair. This result confirms that customer rating, online reviews, land values, and shop rent represent four sufficiently distinct facets of commercial land benefit. This independence supports the analytical decision to model each indicator separately using the XGBoost framework, as the explanatory structures underlying each benefit dimension are likely to differ in meaningful and interpretable ways.

5.1.2. The Spatial Distribution

Because the four indicators differ substantially in their units and measurement scales, direct comparison of raw value maps across dimensions would be misleading. Therefore, the spatial distribution of each indicator was mapped using a within-indicator tercile-based classification, where parcels were assigned to three categories relative to the indicator's own distribution: low (<33rd percentile), middle (33rd–67th percentile), and high (>67th percentile). This normalization ensured that each panel in Figure 5 represented the relative standing of a parcel within each benefit dimension, enabling cross-dimensional comparison of spatial patterns independent of measurement scale.

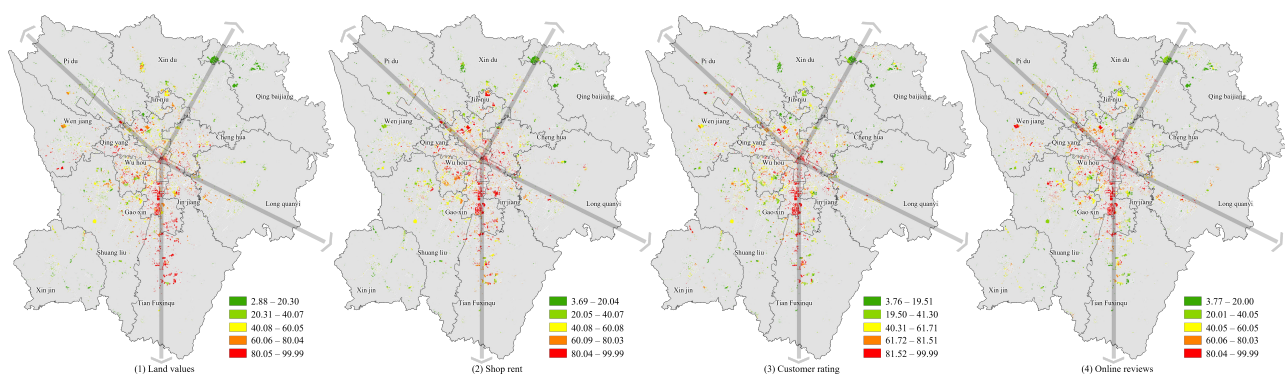


Figure 5. The spatial distribution of 4 benefit variables.

The spatial distributions of all four benefit indicators exhibited a pronounced clustering pattern predominantly within the five legacy urban districts and Gaoxin (the southern segment of the north–south development axis). This spatial configuration was broadly aligned with Chengdu's core urban development strategic priorities in recent years, which were anchored in the legacy urban districts as the foundational core and feature targeted southward expansion along the aforementioned north–south axis.

Furthermore, commercial land parcels in Wenjiang, Shuangliu, Pidu, Xindu, Longquanyi and Tian Fuxinqu that are geographically proximate to the five legacy urban districts and Gaoxin demonstrated relatively higher economic and non-economic benefits, reflecting a commercial development dynamic characterized by centrifugal spillover effects from the urban core. In contrast, commercial land parcels in Qingbaijiang and Xinjin failed to exhibit favorable performance across both the economic and non-economic benefit dimensions.

5.1.3. Differences in Spatial Distribution of the Four Benefit Variables

To assess the spatial alignment of the four commercial benefit dimensions, a pairwise three-way mismatch classification was conducted. For each of the six possible dimension pairs, parcels were first classified as above or below the sample median of each dimension.

The median was adopted as the threshold because it is distribution-agnostic, symmetric, and directly interpretable without assuming normality. Each parcel was then assigned to one of three mutually exclusive categories: (i) A-higher and B-lower (e.g., parcels with above-median land values but below-median customer ratings), (ii) A-lower and B-higher (the reverse direction), or (iii) matched (both dimensions above or both below their respective medians). This ternary scheme allowed the analysis to distinguish between the two opposite directions of spatial divergence, each with potentially distinct underlying predictive factors, rather than collapsing them into a single “mismatched” category.

In total, 70.10% of the two economic indicators fell into the matched category, with 14.8% and 15.0% classified as LV-low and SR-high (low land values and high shop rent, in Figure 6(1)) and LV-high and SR-low (high land values and low shop rent, in Figure 6(2)). LV-low and SR-high zones were predominantly located within the five legacy urban districts, with an average year of completion of 28.1 years. Owing to their land transactions having been conducted in earlier periods, these zones exhibited relatively low overall land prices that are misaligned with their elevated rental levels. In contrast, the majority of LV-high and SR-low commercial land parcels were predominantly newly developed in recent years, with an average year of completion of 6.5 years. These parcels were characterized by inflated land prices, coupled with their generally greater distance from the urban core, lower surrounding built-up intensity and weaker population vitality, thereby resulting in depressed rental levels.

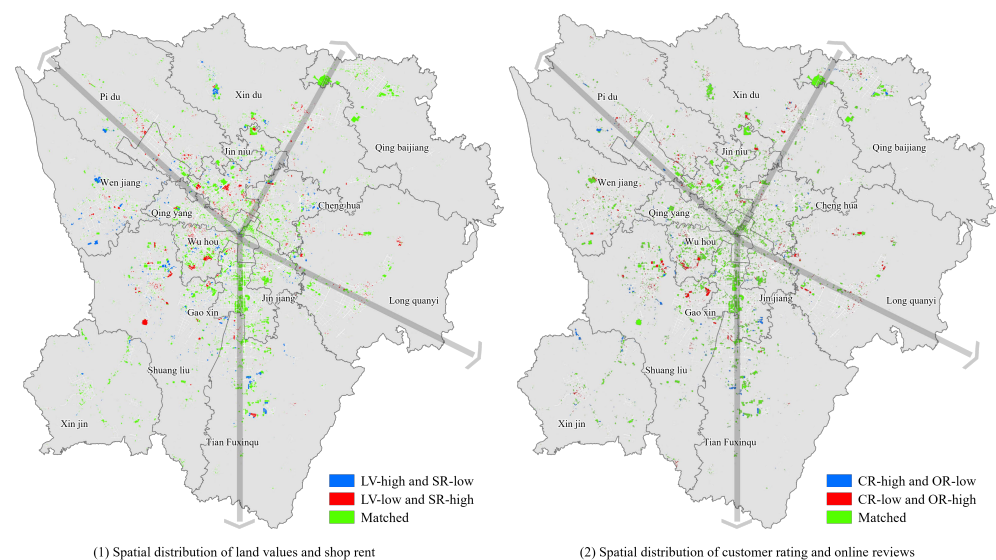


Figure 6. The spatial matching relationship of economic and non-economic indicators.

The customer rating and online reviews pair exhibited a higher level of alignment. Across the 10,766 parcels, 78.1% were classified as matched, indicating that consumer satisfaction and online review volume were broadly consistent across most of Chengdu. The CR-low and OR-high (low customer rating and high online reviews) and CR-high and OR-low (high customer rating and low online reviews) indicators accounted for 11.6% and 10.3%, respectively. Such discrepancies constitute a relatively common occurrence in urban commercial systems. Some commercial spaces that exhibit high population vitality yet receive predominantly negative consumer evaluations may possess indispensable functional or locational advantages, rendering them the only viable option for surrounding residents and consumers. In contrast, commercial spaces that garner consistently positive customer assessments but demonstrate insufficient population vitality may be constrained by factors including limited visitor reception capacity and appointment-only operational models.

Overall, the spatial distribution discrepancies between economic and non-economic indicators were greater than those between any pairwise combination of indicators within the same category. Specifically, mismatch rates of 35.0%, 38.8%, 28.3%, and 32.2% were observed for the land values–customer rating (Figure 7(1)), land values–online reviews (Figure 7(2)), shop rent–online reviews (Figure 7(3)), and shop rent–customer rating pairs (Figure 7(4)), respectively. Interestingly, land parcels where economic indicators outperformed non-economic indicators were predominantly distributed in the southern region of Chengdu (e.g., Gaoxin, Tian Fuxinqu, and Shuangliu). Conversely, land parcels where economic indicators underperformed non-economic indicators were typically located in the northern and inner suburban areas of Chengdu (e.g., Xindu and Qingbaijiang).

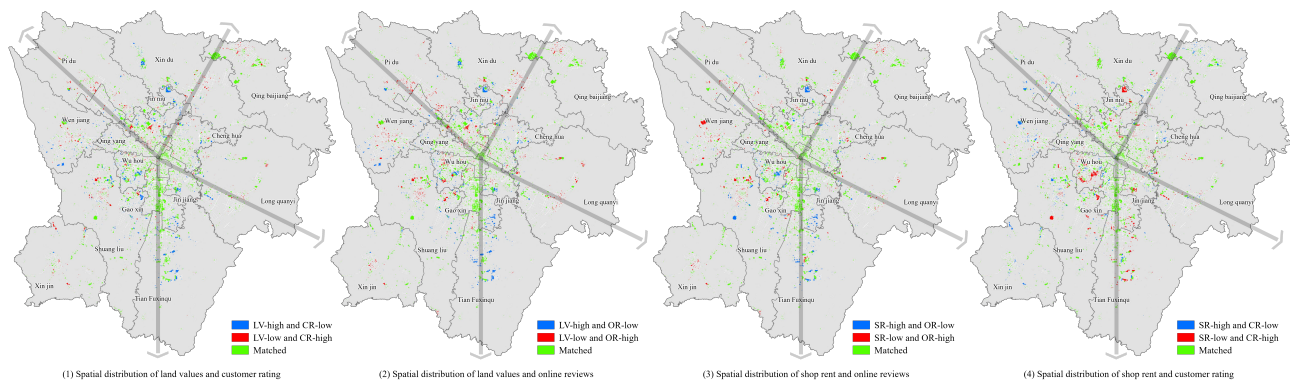


Figure 7. Spatial mismatch distribution of economic versus non-economic benefit indicators across the study area.

5.2. Relative Importance of Independent Variables for Four Benefit Variables

The full 24-variable SHAP importance rankings for all four benefit indicators are presented in Figure 8, and the distribution of top-10 SHAP factors is shown in Figure 9. Additionally, the performance of the model is shown in Table 7.

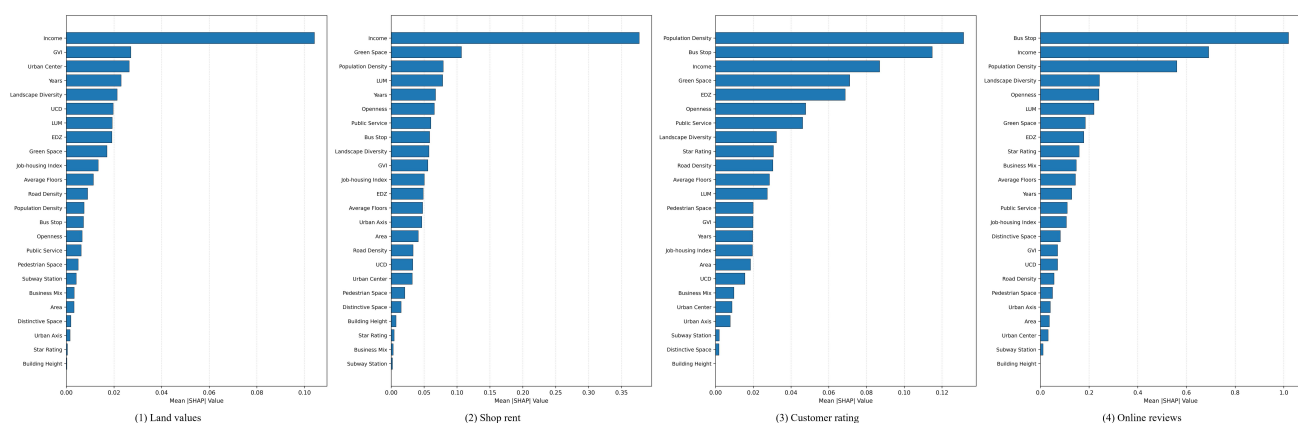


Figure 8. The relative importance of independent variables for 4 benefit variables.

Table 7. The performance of XGBoost-SHAP.

Dependent Variable	Scale	R ² (±SD)	RMSE (±SD)	MAE (±SD)
Land values	Log	0.622 ± 0.023	0.180 ± 0.004	0.123 ± 0.002
Shop rent	Log	0.526 ± 0.016	0.940 ± 0.028	0.605 ± 0.014
Customer rating	Original	0.546 ± 0.022	0.665 ± 0.022	0.298 ± 0.010
Online reviews	Log	0.829 ± 0.007	1.418 ± 0.035	1.025 ± 0.019

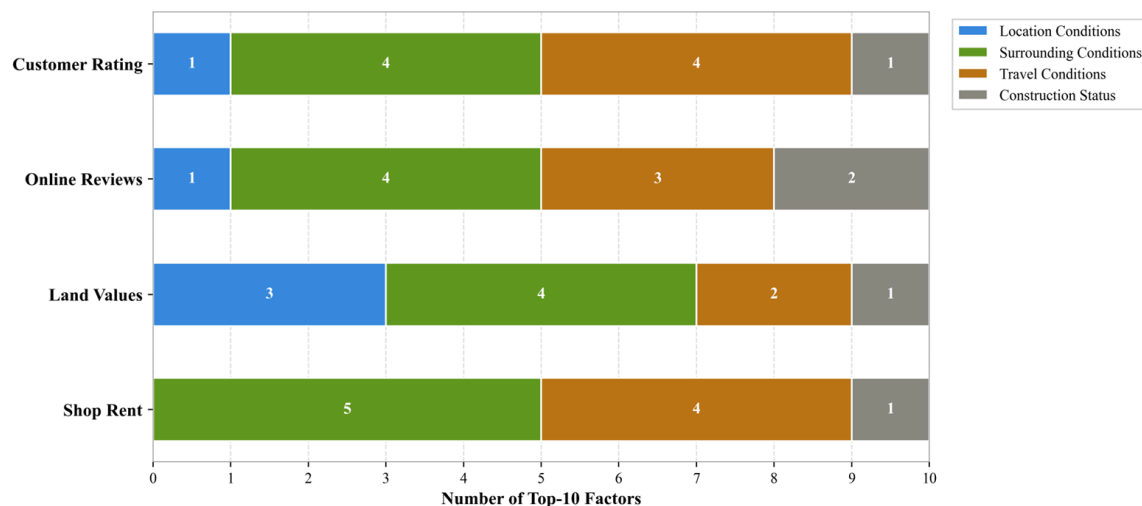


Figure 9. Distribution of top-10 SHAP factors into 4 categories. (The number within each colored segment indicates the count of top-10 ranked features belonging to that category.)

5.2.1. The Analysis of Land Values

For land values, income was the strongest predictor, consistent with the established finding in urban economics that land prices capitalize the purchasing power and socio-economic status of the surrounding population. GVI and landscape diversity, the two perceptual streetscape quality variables, ranked second and fifth, suggesting that the visual greenness and landscape of the surrounding environment emerged as meaningful predictors of land valuation in Chengdu. Urban center, UCD and EDZ ranked third, sixth, and eighth, respectively. This ranking pattern further corroborates the well-established tenet of location theory that urban regional attributes are strongly predictive of variations in land values. Additionally, in academic papers investigating Chengdu's real estate market, it is also documented that street spatial quality and location are critical variables influencing housing prices, rents and other related indicators [11,27].

5.2.2. The Analysis of Shop Rent

For shop rent, income was again the strongest predictor and exhibited the largest SHAP value among all four benefit dimensions. Consistent with land values, variables reflecting streetscape quality, including landscape diversity and GVI and openness, ranked highly in importance, further corroborating the critical role of surrounding spatial quality in determining the economic performance of commercial land uses [23]. However, unlike land values, which were predominantly driven by locational conditions, shop rents appeared to be more strongly influenced by surrounding contextual conditions. Population density, public services, bus stops and LUM all ranked among the top influential factors. This divergence in predictive factors can be understood through the distinct theoretical frameworks.

5.2.3. The Analysis of Online Reviews

The number of bus stops, a proxy for public transit accessibility to commercial spaces, was the dominant predictor for online reviews. These results were observed in papers about the offline vitality of commercial spaces [11] and the online vitality of parks [46]. Population density, income, LUM and green space, which belonged to the surrounding conditions, formed an important basis for the vitality of commercial spaces. Additionally, property-level attributes, including star rating, business mix, average floors, and years, were key predictors of commercial attractiveness.

5.2.4. The Analysis of Customer Rating

For customer rating, most of the top-ranked variables fell into the surrounding conditions and travel conditions. In contrast, online reviewers paid little attention to a commercial venue's inherent business mix, scale and tier. Research has indicated that consumers generally care more about the overall status of its neighborhood [13].

5.3. Effects of Independent Variables

The analysis prioritized the investigation of the four types of independent variables that demonstrated consistently high predictive importance across all four benefit indicators. In the interaction effect plots, the interaction partners for these key variables were automatically selected as the default optimal options.

5.3.1. Effects of Location Conditions

Among location conditions, EDZ was the most important predictor among the four benefit indicators, ranking 8th, 12th, 8th and 4th in terms of importance for land values, shop rent, online reviews and customer rating, respectively. Figure 10 presents SHAP dependence plots for EDZ, with the x-axis showing EDZ status and the y-axis showing the SHAP contribution of EDZ. Each panel is color-coded by its strongest interaction partner: (a) land values $\times$ income ($|r| = 0.34$), (b) shop rent $\times$ bus stops ($|r| = 0.38$), (c) customer rating $\times$ income ($|r| = 0.49$), and (d) online reviews $\times$ income ($|r| = 0.58$). A clear and consistent divergence emerges between economic and non-economic outcome variables.

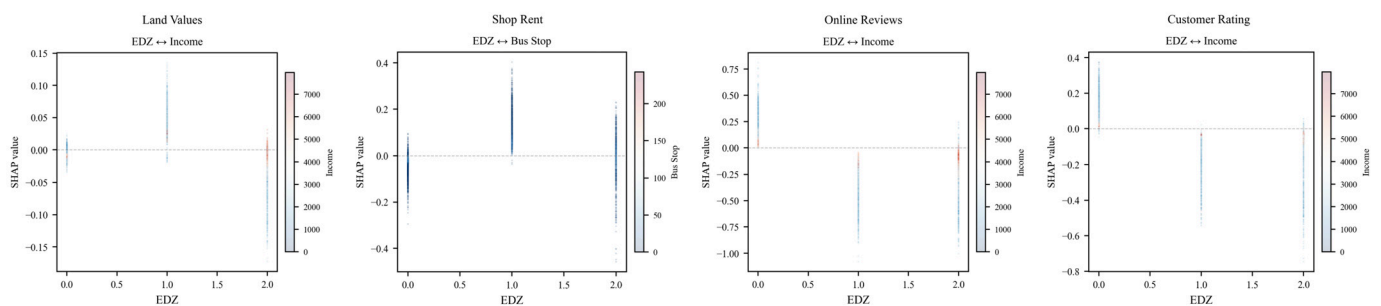


Figure 10. Interaction effects of EDZ on 4 benefit variables.

The SHAP dependence plots for land values and shop rent both exhibited an inverted U-shaped pattern. Average SHAP values were -0.008 and -0.032 for land values and shop rent when $EDZ = 0$ and increased to 0.051 and 0.111 when $EDZ = 1$, while the corresponding SHAP values deteriorated to -0.038 and -0.011 when $EDZ = 2$. In contrast, customer rating and online reviews show a monotonically decreasing trend. Furthermore, our findings further demonstrate that lower-income groups exhibited a more pronounced amplifying effect on both the positive and negative effects of EDZ. In contrast, low-density bus stops primarily amplified the negative effects of EDZ.

In general, EDZ, representing high-tech industrial agglomeration, is expected to show a stronger positive predictive association with land values, which is corroborated by the transition from $EDZ = 0$ to $EDZ = 1$. However, the delineation rules for EDZ in Chengdu specify that district-level EDZ is predominantly composed of commercial centers and business districts, while municipal-level EDZ is mainly dominated by industrial parks, manufacturing zones and biomedical R&D centers. Given the inherent incompatibility between industrial/manufacturing/biomedical R&D activities and commercial spaces, this is consistent with the observed decline in economic indicators of commercial land parcels at $EDZ = 2$. Furthermore, the extant literature on China's industrial agglomeration and development zones indicates that EDZ-driven regional development models tend to

prioritize large-scale landmark projects at the expense of street-level commercial diversity and pedestrian-oriented amenities, thereby reducing customer visit rates and objective consumer evaluations.

5.3.2. Effects of Surrounding Conditions

Within the category of surrounding conditions, income and green space emerged as the two most important predictors across all four benefit indicators. Specifically, income occupied the first, first, second and third positions in the relative importance ranking for land values, shop rent, online reviews and customer rating, respectively, whereas green space ranked ninth, second, seventh and fifth, respectively, among all covariates included in the model.

Figure 11 shows the effects of income and its interaction effects with landscape diversity, openness and EDZ. Overall, income exerted pronounced positive effects on all four benefit indicators. In particular, the positive effects of income on land values, shop rent and online reviews were distinct, robust and nearly linear. In contrast, the positive predictive association of income with customer rating was substantially weaker, indicating that consumers' evaluations of commercial spaces showed a notably weaker association with their income levels, which was consistent with our previous analyses. Furthermore, indicators reflecting spatial quality, such as landscape diversity and openness, positively amplified the effects of income, consistent with the existing literature [15], while the interaction effects of EDZ aligned with the findings presented in Figure 10.

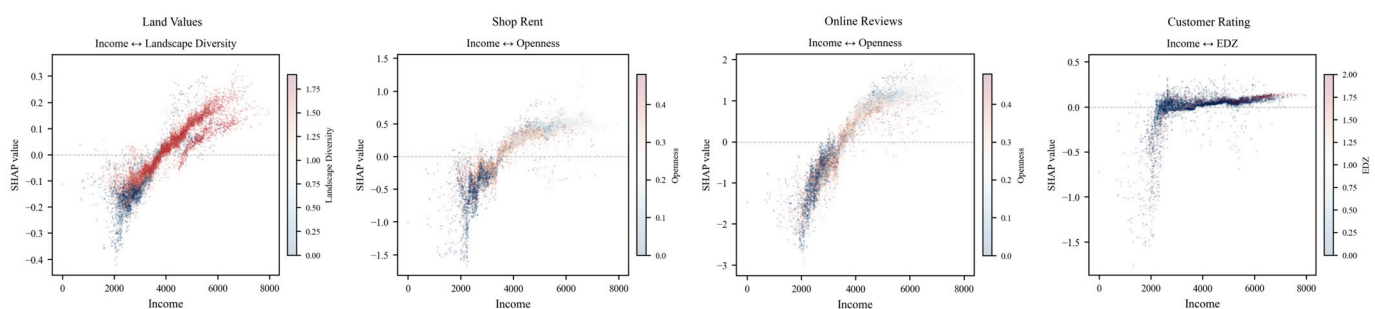


Figure 11. Interaction effects of income on 4 benefit variables.

A higher concentration of high-income populations surrounding commercial land parcels indicates greater purchasing power and consumer demand, which is associated with higher vitality indicators such as online reviews, land values, and shop rents, and is correspondingly correlated with a larger proportion of high-income populations—consistent with a mutually reinforcing spatial pattern between economic vitality and neighborhood socio-economic composition, and vice versa. Furthermore, the overall spatial quality of commercial parcel locations shows a synergistic positive predictive association, particularly as high-income groups exhibit significantly greater sensitivity and preference for spatial quality [26].

Figure 12 shows the effects of green space and its interaction effects with income and bus stops. The SHAP dependence plots for green space revealed a consistent positive gradient across all four benefit dimensions, with interaction patterns falling into two distinct regimes based on the strongest interaction partner: an income-conditioned regime for market valuation dimensions (interacting with income, $|r| = 0.33\text{--}0.42$) and an accessibility-mediated regime for consumer experience dimensions (interacting with bus stops, $|r| = 0.41$ for both).

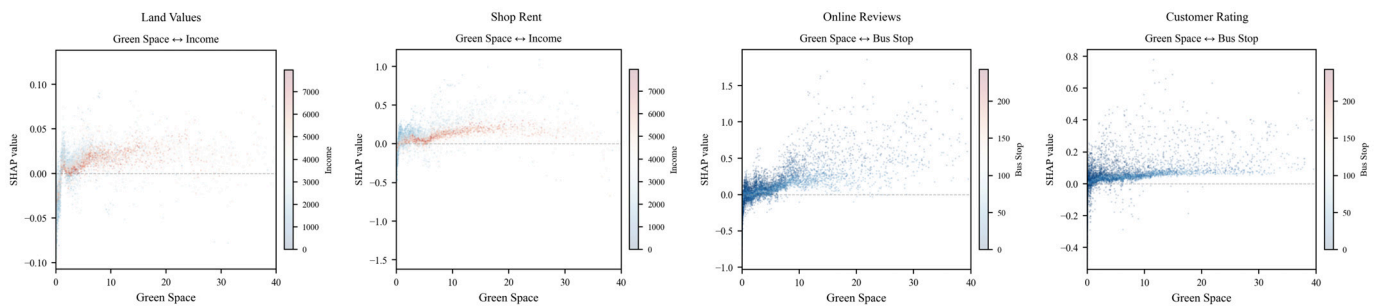


Figure 12. Interaction effects of green space on 4 benefit variables.

The largest incremental gain in SHAP contribution occurred between the first and second quartiles—that is, the move from near-zero green space ($\leq 0.30\%$ in the 500 m buffer) to minimal provision. This pattern suggests that, from a planning perspective, the highest return on investment lies not in maximizing green space but in ensuring that no commercial parcel falls below a basic minimum threshold. This finding resonates with the urban planning literature on green space equity [33], which posits that the marginal social benefit of park provision is greatest at the base of the distribution.

High-income populations consistently amplified the positive effects of green space on economic factors, consistent with a positive predictive association between socio-economic attributes and spatial quality surrounding commercial land parcels. Conversely, an insufficient number of bus stops weakened the positive effect of green space on online reviews and customer ratings. This pattern is consistent with the expectation that inadequate public transit accessibility may limit both the likelihood and willingness of people to access commercial spaces, leading to lower customer evaluation outcomes.

5.3.3. Effects of Travel Conditions

Within the category of travel conditions, landscape diversity emerged as the most important predictor and exhibited fluctuating positive predictive associations across all four benefit indicators in Figure 13, which occupied the 5th, 9th, 4th and 14th positions. Landscape diversity was accompanied by a pronounced interaction with other spatial quality indicators, including GVI and pedestrian space.

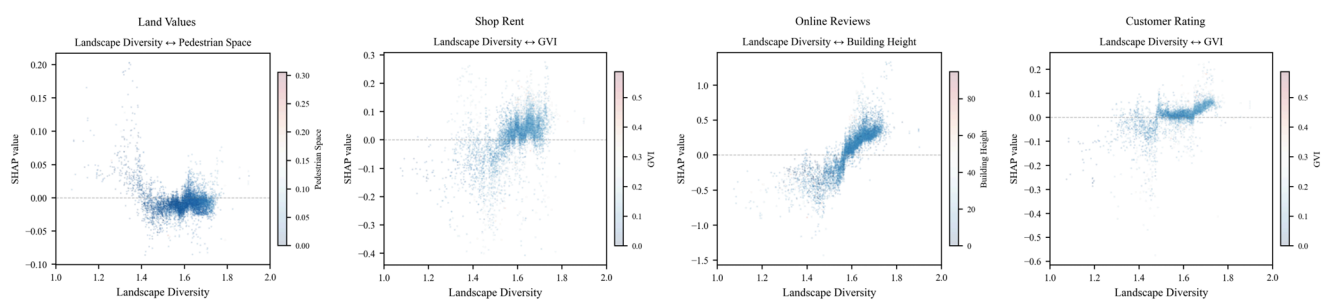


Figure 13. Interaction effects of landscape diversity on 4 benefit variables.

Higher landscape diversity was predictively associated with more favorable outcomes across all four commercial land value dimensions, though the shape and magnitude of this association varied by outcome type. For shop rent, online reviews, and customer rating, a threshold structure was observed: SHAP values shifted from negative to positive at approximately 1.5–1.6. Below this level, limited landscape diversity appeared to be associated with reduced consumer engagement and willingness-to-pay, and once diversity exceeded this threshold, the association stabilized or moderately increased. This pattern is

consistent with findings from urban design research that mixed visual environments attract more visitor traffic and longer dwell times [26,34].

For online reviews and shop rent, the co-occurrence of a higher GVI with higher landscape diversity suggests that green visual quality amplifies the positive association of mixed land use with commercial vitality, consistent with evidence that vegetation enhances the perceived attractiveness of diverse urban settings [47]. For land values, pedestrian space emerged as the interaction partner, consistent with the hypothesis that walkable environments facilitate access to diverse land uses and thereby influence parcel-level economic valuation [28].

5.3.4. Effects of Construction Status

Within the category of construction status, years emerged as the most important predictor across all four benefit indicators, which occupied the 4th, 5th, 12th and 17th positions in Figure 14. The results revealed divergent relationships between years (building age) and commercial land value outcomes, with the direction of the predictive association depending on the outcome measure under consideration.

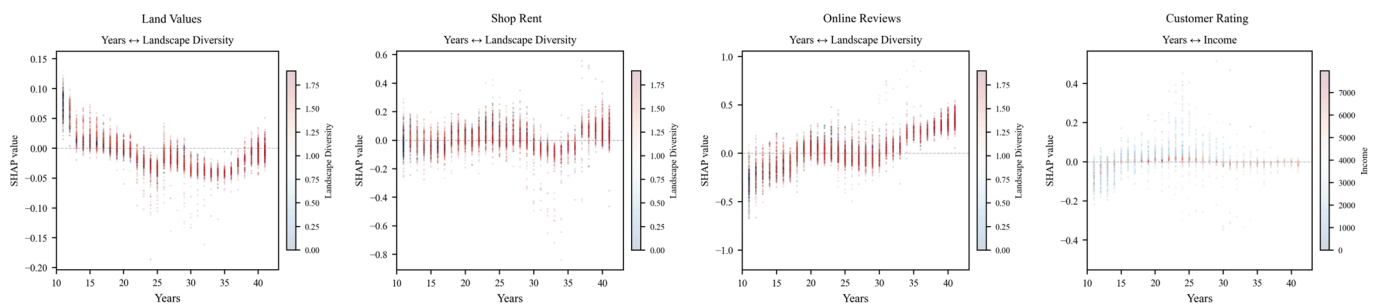


Figure 14. Interaction effects of years on 4 benefit variables.

The monotonic decline in SHAP values with increasing years for land values is consistent with standard real estate depreciation theory: older buildings are typically associated with lower capital values due to physical deterioration, functional obsolescence, and higher maintenance costs [9]. The positive SHAP values observed exclusively at Q4 for shop rent and online reviews suggest that older commercial buildings in Chengdu may benefit from established commercial reputations, distinctive architectural character, and prime locations in historic commercial districts, consistent with findings that building age can positively contribute to retail rents when it serves as a proxy for location centrality and brand heritage. The intermediate-age decline (Q3, -0.038) may reflect a “middle-age trap” in which commercial buildings are neither new enough to attract modern retail tenants nor old enough to possess heritage appeal.

Landscape diversity partially offset the depreciation effect of years on land values and strengthened the appreciation influence of years on shop rent and online reviews. Buildings in diverse settings may retain value better than age alone would predict. From a planning perspective, this finding implies that targeted urban renewal in older, high-diversity neighborhoods may unlock latent economic potential and the number of visitors. The substitution of landscape diversity by income as the dominant interaction partner for customer rating ($|r| = 0.211$) is informative. Consumer satisfaction, unlike economic values or online engagement, appears to be more strongly shaped by the economic profile of the surrounding population than by the physical diversity of the built environment. A hump-shaped relationship can be observed, with the SHAP value peaking at Q2 (years ≈ 20.6 , income ≈ 3763) and declining monotonically thereafter, suggesting diminishing marginal returns to neighborhood affluence above a specific threshold. This finding resonates with

the work of Zhang et al. [48], who documented that place-based policies in China often exhibit nonlinear effects, with benefits concentrated in middle-income neighborhoods rather than the most affluent ones.

6. Conclusions

This study constructs a multidimensional benefit evaluation framework for urban commercial land, encompassing four indicators, and applies XGBoost-SHAP methods to analyze their predictive associations and spatial patterns in Chengdu. The principal findings are summarized as follows:

- (1) The four benefit indicators capture substantially independent dimensions of commercial land performance. All pairwise Pearson correlations are low, confirming their distinctiveness and justifying separate XGBoost modeling. This independence suggests that future research may incorporate additional benefit perspectives to form a more comprehensive evaluation system.
- (2) Spatially, high-value parcels concentrated within Chengdu's five legacy urban districts, Gaoxin, and the southern urban axis reflect the implementation outcomes of Chengdu's development strategies and urban-center spillover effects. However, spatial mismatch analysis reveals that inter-category divergence substantially exceeds intra-category variation: 28.3–38.8% of parcels are mismatched, with southern districts overperforming on economic dimensions and northern districts showing the reverse pattern.
- (3) SHAP rankings reveal functional specialization across benefit dimensions. Land values are predicted primarily by locational attributes and streetscape quality, consistent with location theory; shop rent, while also predicted by income, is more strongly associated with surrounding contextual conditions, reflecting distinct theoretical explanations underlying land pricing versus rental determination.
- (4) For non-economic indicators, online reviews integrate both neighborhood conditions and property-level attributes, whereas customer rating is almost exclusively associated with surrounding and travel conditions, with venues' inherent characteristics receiving little attention, confirming that consumers evaluate commercial experiences through neighborhood context rather than establishment attributes.
- (5) Among location conditions, EDZ exhibits a consistent divergence: an inverted-U pattern for economic outcomes versus a monotonic decline for non-economic outcomes, attributable to the incompatibility between municipal-level industrial parks and commercial vitality. Income is the most influential variable across all dimensions, with near-linear positive associations on three of four outcomes but substantially weaker influence on customer rating. Green space shows a consistent positive gradient, with the largest marginal SHAP gain at the transition from near-zero to minimal provision—implying that ensuring basic green coverage yields higher planning returns than maximizing provision in already-green areas.
- (6) Landscape diversity displays a threshold structure at approximately 1.5–1.6, below which limited visual diversity is associated with reduced consumer engagement. Building age exhibits divergent patterns: monotonic depreciation for land values, a heritage premium for the oldest cohort in shop rent and online reviews, and a hump-shaped trajectory for customer rating peaking at approximately 20 years.
- (7) Landscape diversity and spatial quality indicators systematically moderate primary driver effects across all analyses, partially offsetting building-age depreciation on land values while enhancing the vintage premium for consumer outcomes. Income, GVI, and pedestrian space positively amplified landscape diversity's benefits, whereas inadequate transit accessibility weakened green space's positive effects. This consis-

tent moderating role positions the surrounding built environment as a critical context modifier—revealing how locational, socio-economic, and property-level predictors are related to economic and non-economic outcomes—rather than merely an independent predictor. The substitution of landscape diversity by income as the dominant interaction partner for customer rating further indicates that consumer satisfaction is more strongly associated with neighborhood socio-economic context than physical built-environment diversity.

These findings carry differentiated implications for three stakeholder groups. For urban planners and policymakers, the spatial mismatch between economic and non-economic benefit indicators (28.3–38.8% of parcels), coupled with the consistent north–south divergence pattern, underscores the need for sub-regionally differentiated rather than uniform commercial land policies: southern districts where economic indicators systematically outperform non-economic ones may benefit from consumer-experience-oriented interventions, whereas northern districts where the reverse pattern holds may require economic revitalization strategies. For developers and investors, the empirical distinctiveness of the four benefit dimensions indicates that multi-indicator due diligence—incorporating consumer engagement metrics alongside traditional price indicators—yields a more complete assessment of commercial land potential than price-only evaluation. For commercial operators, the consistent moderating role of spatial quality indicators on economic outcomes offers actionable guidance for location decisions: parcels in visually diverse, green-rich, and pedestrian-accessible neighborhoods systematically outperform otherwise comparable parcels, particularly within higher-income catchments.

Several limitations warrant acknowledgment. First, the cross-sectional design captures associations in a single temporal slice and cannot establish causal directionality or track how mismatches evolve over time. Second, the reliance on online platform data (Dianping reviews and POI-derived ratings) introduces potential selection bias toward digitally active consumers and may underrepresent demographic segments with lower platform adoption rates. Third, while Chengdu's polycentric structure offers a valuable contrast to monocentric cities typically studied in the Western literature, the generalizability of specific threshold estimates to cities with different institutional contexts, development stages, or market structures remains an open question requiring comparative investigation. Furthermore, within the study area itself, the 14 retained districts span heterogeneous development stages, ranging from post-industrial service economies to manufacturing-oriented growth poles. Although all districts were retained to avoid selection bias and each exceeds a 50% tertiary GDP share, inter-district variation in commercial maturity and consumer behavior may affect the consistency of benefit indicator interpretation across sub-regions, warranting district-stratified robustness testing in future work. Additionally, although indicators like EDZ and UCD, representing policies and strategic intentions, have been selected for research, the direct impact of policies has not been fully examined, which could be a key direction for future research. Our framework excludes certain potentially relevant factors—notably, ownership structure, lease terms, and management quality—which future research incorporating primary survey data could productively address. Fourth, although XGBoost modeling substantially reduces residual spatial autocorrelation (53.7–76.8% reduction), Moran's I tests confirm that statistically significant spatial dependence persists in the residuals across all four benefit indicators ($p < 0.001$). This suggests that while the models capture a large portion of spatial structure, some unexplained geographic clustering remains, which future applications may address by incorporating explicit spatial structuring (e.g., spatial lags or geographically weighted regression) into the XGBoost-SHAP framework. Furthermore, while random cross-validation yields robust in-sample predictive performance ($R^2 = 0.526$ – 0.829 across the four dependent variables), the model's

out-of-sample spatial generalization is substantially weaker; for instance, the land value R^2 drops from 0.622 to 0.050 under spatial cross-validation, indicating that the learned predictive associations do not generalize well to geographically held-out test sets.

Finally, the modeling framework excludes several factors that may contribute to the observed benefit mismatches. Beyond parcel-level attributes such as ownership structure, lease terms, and property-rights duration, market-level mechanisms—including advertising expenditures, social media influencer effects (viral online attention), and platform-specific recommendation algorithms—were not incorporated owing to data availability constraints. These omitted factors may partially account for residual divergences between economic and non-economic indicators.

Author Contributions: Conceptualization, Y.J. and S.S.; methodology, Y.J. and P.T.; software, Y.J., Y.L. and G.Z.; formal analysis, P.T. and S.S.; validation, B.X.; investigation G.Z.; resources Y.J., P.T., S.S. and G.Z. Data curation, Y.J. and X.Z.; Writing—original draft, Y.J., P.T. and Y.L.; Writing—review & editing, Y.J. and S.S.; Supervision, P.T. and S.S.; Project administration, S.S.; Funding acquisition, B.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China [Grant No. 52378077].

Data Availability Statement: If necessary, data can be provided, but an agreement must be signed.

Conflicts of Interest: Author Gaomeiyuan Zheng was employed by the company SHUDAO Investment Group Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LUM	Land-use mix
GVI	Green view index
XGBoost	eXtreme Gradient Boosting
SHAP	SHapley Additive exPlanations
UCD	Urban Commercial District
EDZ	Economic Development Zone
POI	Point of Interest
ALE	Accumulated Local Effects
LIME	Local Interpretable Model-agnostic Explanations
CV	Cross-validation

References

1. Xu, T.; Zhang, M.; Aditjandra, P.T. The impact of urban rail transit on commercial property value: New evidence from Wuhan, China. *Transp. Res. Part A Policy Pract.* **2016**, *91*, 223–235. [[CrossRef](#)]
2. Colaço, R.; de Abreu e Silva, J. Commercial land use change and growth processes—An assessment of retail location in Lisbon, Portugal, 1995–2020. *J. Urban Manag.* **2024**, *13*, 157–170. [[CrossRef](#)]
3. Ecker, M.D.; Isakson, H.R. A unified convex–concave model of urban land values. *Reg. Sci. Urban Econ.* **2005**, *35*, 265–277. [[CrossRef](#)]
4. Nelson, A.C. Transit Stations and Commercial Property Values: A Case Study with Policy and Land-Use Implications. *J. Public Transp.* **1999**, *2*, 77–95. [[CrossRef](#)]
5. Alonso, W. *Location and Land Use: Toward a General Theory of Land Rent*; Harvard University Press: Cambridge, MA, USA, 1964.
6. Davis, M.A.; Oliner, S.D.; Pinto, E.J.; Bokka, S. Residential land values in the Washington, DC metro area: New insights from big data. *Reg. Sci. Urban Econ.* **2017**, *66*, 224–246. [[CrossRef](#)]
7. Byun, J.; Jang, K. The effects of subway operation for commercial land values: A case study in Daejeon, South Korea. *Res. Transp. Econ.* **2024**, *106*, 101462. [[CrossRef](#)]

8. Liu, X.; Tong, D.; Huang, J.; Zheng, W.; Kong, M.; Zhou, G. What matters in the e-commerce era? Modelling and mapping shop rents in Guangzhou, China. *Land Use Policy* **2022**, *123*, 106430. [[CrossRef](#)]
9. Peng, C.; Yang, S.; Zhang, P.; Hu, S. Exploring nonlinear and interaction effects of TOD on housing rents using XGBoost. *Cities* **2025**, *158*, 105728. [[CrossRef](#)]
10. Wang, L.; Yuan, F.; Duan, X. How high-speed rail service development influenced commercial land market dynamics: A case study of Jiangsu province, China. *J. Transp. Geogr.* **2018**, *72*, 248–257. [[CrossRef](#)]
11. Zhang, W.; Li, Y.; Zheng, C. The distribution characteristics and driving mechanism of vacant land in Chengdu, China: A perspective of urban shrinkage and expansion. *Land Use Policy* **2023**, *132*, 106812. [[CrossRef](#)]
12. Gebru, M.A.; Abebe, T.; Zewale, H.L.; Enyew, F.F.; Asnakew, M.B.; Nega, W. Drivers and socio-economic implications of vacant urban land: A case study of Debre Markos City, Ethiopia. *Cities* **2026**, *170*, 106623. [[CrossRef](#)]
13. Wang, T.; Wang, Y.; Zhao, X.; Fu, X. Spatial distribution pattern of the customer count and satisfaction of commercial facilities based on social network review data in Beijing, China. *Comput. Environ. Urban Syst.* **2018**, *71*, 88–97. [[CrossRef](#)]
14. Mao, H.; Fan, X.; Guan, J.; Chen, Y.-C.; Su, H.; Shi, W.; Zhao, Y.; Wang, Y.; Xu, C. Customer attractiveness evaluation and classification of urban commercial centers by crowd intelligence. *Comput. Hum. Behav.* **2019**, *100*, 218–230. [[CrossRef](#)]
15. Akaabre, P.B. Traditional leasehold of land for residential and commercial use in Ghana: Structure and practices from the Golden Stool. *Land Use Policy* **2023**, *131*, 106747. [[CrossRef](#)]
16. Li, Z.; Hu, L.; Lin, A.; Chen, J.; Xu, Y. The greener, the richer, the happier?—Spatial distribution and coupling analysis of urban green space and residents' emotion based on social media data. *Ecol. Indic.* **2025**, *177*, 113754. [[CrossRef](#)]
17. Guo, C.; Yang, Y. A multi-modal social media data analysis framework: Exploring the complex relationships among urban environment, public activity, and public perception—A case study of Xi'an, China. *Ecol. Indic.* **2025**, *171*, 113118. [[CrossRef](#)]
18. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
19. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 4–9 December 2017; Volume 30.
20. Colwell, P.F.; Munneke, H.J. The Structure of Urban Land Prices. *J. Urban Econ.* **1997**, *41*, 321–336. [[CrossRef](#)]
21. McDonald, J.F.; McMillen, D.P. *Urban Economics and Real Estate: Theory and Policy*; John Wiley & Sons: Hoboken, NJ, USA, 2010.
22. Liu, Y.; Zhang, C.; Zhao, S. Supply-Demand Mismatch of Urban Commercial Land and Its Impact Mechanism in Gansu Province Based on an Explainable Machine Learning Model. *Land* **2026**, *15*, 351. [[CrossRef](#)]
23. Wang, H.; Wu, K.; Feng, Z.; Zhao, H.; Ai, H.; Meng, C. Evaluation of Urban Commercial Land Use Intensification Based on Land Parcels: Taking Wuxi City as an Example. *Land* **2023**, *12*, 1608. [[CrossRef](#)]
24. Li, L.; Du, Q.; Ren, F.; Huang, L.; Voda, M.; Ning, P. Geolocated social media data for measuring park visitation in Shenzhen, China. *Urban For. Urban Green.* **2023**, *88*, 128069. [[CrossRef](#)]
25. Li, X.; Pang, C. A spatial visual quality evaluation method for an urban commercial pedestrian street based on streetscape images—Taking Tianjin Binjiang road as an example. *Sustainability* **2024**, *16*, 1139. [[CrossRef](#)]
26. Huang, G.; Yu, Y.; Lyu, M.; Sun, D.; Dewanker, B.; Gao, W. Impact of physical features on visual walkability perception in urban commercial streets by using street-view images and deep learning. *Buildings* **2024**, *15*, 113. [[CrossRef](#)]
27. Gao, Y.; Cui, X.; Sun, X.; Feng, H.; Liang, P.; Yu, B. Exploring Transit-Oriented Development (TOD) frameworks for commercial agglomeration around rail stations: A case study of Chengdu, China. *J. Asian Archit. Build. Eng.* **2025**, 1–20. [[CrossRef](#)]
28. Ewing, R.; Cervero, R. Travel and the built environment: A meta-analysis. *J. Am. Plan. Assoc.* **2010**, *76*, 265–294. [[CrossRef](#)]
29. Cervero, R.; Kockelman, K. Travel demand and the 3Ds: Density, diversity, and design. *Transp. Res. Part D Transp. Environ.* **1997**, *2*, 199–219. [[CrossRef](#)]
30. Christaller, W. *Die Zentralen Orte in Süddeutschland: Eine Ökonomisch-Geographische Untersuchung Über Die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen Mit Städtischen Funktionen*; Wissenschaftliche Buchgesellschaft: Darmstadt, Germany, 1933.
31. Yen, B.T.; Feng, C.-M.; Lee, T.-C. Transit-oriented development strategy in Taiwan: An application of land value capture. *Asian Transp. Stud.* **2023**, *9*, 100094. [[CrossRef](#)]
32. Hu, S.; Chen, M.; Jiang, Y.; Sun, W.; Xiong, C. Examining factors associated with bike-and-ride (BnR) activities around metro stations in large-scale dockless bikesharing systems. *J. Transp. Geogr.* **2022**, *98*, 103271. [[CrossRef](#)]
33. Kim, S.; Woo, A. Streetscape and business survival: Examining the impact of walkable environments on the survival of restaurant businesses in commercial areas based on street view images. *J. Transp. Geogr.* **2022**, *105*, 103480. [[CrossRef](#)]
34. Jiang, Y.; Han, Y.; Liu, M.; Ye, Y. Street vitality and built environment features: A data-informed approach from fourteen Chinese cities. *Sustain. Cities Soc.* **2022**, *79*, 103724. [[CrossRef](#)]
35. Wang, J.; Tan, W.; Huang, J. Extending TOD through the interrelationship between transport and land use: A case study of Beijing. *Land Use Policy* **2024**, *144*, 107257. [[CrossRef](#)]

36. Lin, Y.; Geertman, S. Can social media play a role in urban planning? A literature review. In Proceedings of the International Conference on Computers in Urban Planning and Urban Management, Wuhan, China, 8–12 July 2019; pp. 69–84.
37. Nieweler, S. Path Dependency and the Composition of Land Use, Built Form and Establishment Diversity in the Downtowns of Northwest British Columbia. Ph.D. Thesis, Simon Fraser University, Burnaby, BC, Canada, 2024.
38. Liu, X.; Chen, X.; Tian, M.; De Vos, J. Effects of buffer size on associations between the built environment and metro ridership: A machine learning-based sensitive analysis. *J. Transp. Geogr.* **2023**, *113*, 103730. [[CrossRef](#)]
39. Rosen, S. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *J. Political Econ.* **1974**, *82*, 34–55. [[CrossRef](#)] [[PubMed](#)]
40. Brunsdon, C.; Fotheringham, S.; Charlton, M. Geographically weighted regression. *J. R. Stat. Soc. Ser. D (Stat.)* **1998**, *47*, 431–443. [[CrossRef](#)]
41. Jiao, J.; An, R.; Du, D.; Zhu, M. Non-linear and heterogeneous relationship between proximity to high-speed rail stations and land value in China: Analysis using XGBoost-SHAP modelling. *Transp. Res. Part A Policy Pract.* **2025**, *196*, 104486. [[CrossRef](#)]
42. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.
43. Apley, D.W.; Zhu, J. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2020**, *82*, 1059–1086. [[CrossRef](#)]
44. Hu, Y.; Wu, C.; Meadows, M.E.; Feng, M. Pixel level spatial variability modeling using SHAP reveals the relative importance of factors influencing LST. *Environ. Monit. Assess.* **2023**, *195*, 407. [[CrossRef](#)] [[PubMed](#)]
45. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [[CrossRef](#)] [[PubMed](#)]
46. Wilkins, E.J.; Wood, S.A.; Smith, J.W. Uses and Limitations of Social Media to Inform Visitor Use Management in Parks and Protected Areas: A Systematic Review. *Environ. Manag.* **2020**, *67*, 120–132. [[CrossRef](#)] [[PubMed](#)]
47. Wang, R.; Browning, M.H.E.M.; Qin, X.; He, J.; Wu, W.; Yao, Y.; Liu, Y. Visible green space predicts emotion: Evidence from social media and street view data. *Appl. Geogr.* **2022**, *148*, 102803. [[CrossRef](#)]
48. Zhang, N.; Jiang, Y.; Tang, J.; Zhou, Y.; Yu, R. Evaluation of the Development Benefits and Fairness of TOD in Megacities: A Case Study of Chengdu. *Urban Transp. China* **2024**, *22*, 17–30. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.