Threshold-Sensitive Business Analytics for Urban Retail Density: Interpretable Machine Learning Evidence from Multi-Source Geospatial Data

Farhan Lim[1], Nur Aisyah Rahman[2,*] Hui Min Tan[3]

[1]Department of Business Analytics, Faculty of Business and Management, UCSI University, Kuala Lumpur 56000, Malaysia

[2]Department of Decision Science, Faculty of Business and Finance, Universiti Tunku Abdul Rahman, Kampar 31900, Malaysia

[3]Department of Information Systems, Faculty of Computing and Informatics, Multimedia University, Cyberjaya 63100, Malaysia

*Email: aisyah.rahman@utar.edu.my (Corresponding Author)

Abstract

Urban retail density is increasingly shaped by the interaction between local demand, facility networks, competitive pressure, street-level perception, and accessibility. Yet many business analytics models used for retail site assessment still assume linear responses and therefore understate the threshold behavior that governs where retail clusters become viable, saturated, or strategically fragile. This study develops a threshold-sensitive business analytics framework for urban retail density using interpretable machine learning and multi-source geospatial data. The analytical design integrates retail points of interest, population mobility, road-network accessibility, urban service facilities, competitive context, and street-view perception indicators into a 500-meter grid-level dataset. XGBoost, random forest, and ordinary least squares models are compared, and SHAP-based feature attribution, partial dependence analysis, and threshold interpretation are used to translate model outputs into business and planning insights. The results show that nonlinear machine learning improves predictive performance over linear baselines, with the strongest gains observed for small-format, light-asset retail. Urban function and competition explain the largest share of light-asset retail density, while human perception, accessibility, and facility synergy are more important for capital-intensive retail. Several variables display clear threshold regimes: moderate general-market density supports small-format clustering, high aesthetic perception is required before large-format density increases, and excessive mobility may weaken the stability needed for capital-intensive sites. The findings contribute to business data analytics by demonstrating how interpretable machine learning can move retail location analysis from correlation ranking toward threshold-aware decision support.

Keywords: Urban retail density; business analytics; interpretable machine learning; XGBoost; SHAP; street-view perception; geospatial data; threshold effects

Article History:

Received: January 13, 2025

Revised: March 18, 2025

Accepted: May 21, 2025

Available Online: June 30, 2025

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2025.030205

Threshold-Sensitive Business Analytics for Urban Retail Density: Interpretable Machine Learning Evidence from Multi-Source Geospatial Data

1. Introduction

Retail density is a visible expression of urban commercial vitality. Convenience stores, small supermarkets, large supermarkets, shopping malls, and service-oriented retail outlets do not simply occupy available space; they reveal how businesses read local demand, interpret surrounding facilities, respond to competitors, and evaluate the quality of the street environment. For platform-based commerce and online-to-offline retail, digital channels have changed consumer search and payment behavior, but physical proximity remains critical for immediate consumption, daily service access, neighborhood resilience, and last-mile fulfillment. The central business question is therefore not whether location still matters, but how location matters under increasingly complex urban conditions. This perspective aligns with the wider development of management analytics (Lu,2021). Semantic urban scene datasets provide a foundation for street-level perception extraction (Cordts,2016).

Traditional retail location studies have long relied on demographic indicators, transport accessibility, land value, and distance to competitors. These variables remain useful, but they are often analyzed through linear models that describe average relationships. Such models are attractive because they are easy to interpret, yet they can miss a core property of urban retail systems: many location factors are valuable only within specific ranges. A moderate density of nearby general markets may create agglomeration and attract pedestrians, whereas excessive density may intensify competition and depress additional entry. A visually pleasant streetscape may matter little for routine convenience shopping until it crosses a quality threshold that supports destination retail and leisure consumption. These examples show why threshold-sensitive analytics is important for both business strategy and urban planning. Similar data-source expansion has been documented in urban street-view analytics (Biljecki and Ito,2021). E-shopping complementarity research supports the continued role of local retail environments (Weltevreden,2007).

The rapid development of multi-source geospatial data provides an opportunity to reframe retail location analysis. Retail points of interest identify store locations and competitive local structures. Population heat maps capture temporal mobility and latent demand. Street-view images provide proxies for perceived safety, beauty, greenness, liveliness, and affluence. Open road networks allow analysts to measure integration, route choice, and local connectivity. When these data streams are organized at a fine spatial scale, they enable a business analytics perspective that treats retail density as the outcome of multiple interacting market signals rather than a single factor such as population size or distance to transit. Comparable retail screening work also shows the value of machine learning in location decisions (Lu,2024). Interpretable models show that actionable risk scores can be more useful than opaque accuracy gains (Caruana,2015).

However, the use of machine learning in retail location analysis creates its own challenge. Predictive accuracy alone is not enough for business decision-making. Retail managers, landlords, commercial planners, and city officials need to understand which variables matter, how they matter, and when their effects change direction. Black-box predictions may identify high-density zones, but they do not explain whether a site is attractive because of facility synergy, competitive spillovers, mobility, perceived quality, or network accessibility. The practical value of business analytics depends on the conversion of model outputs into thresholds, risk warnings, and managerial actions. The reliance on volunteered and platform-derived geographic data follows a broader turn toward citizen and platform sensing (Goodchild,2007). Large-scale street-scene datasets support the feasibility of urban visual measurement (Neuhold,2017).

This article develops an interpretable and threshold-sensitive business analytics framework for urban retail density. The proposed framework is directly aligned with recent geospatial retail research showing that light-

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2025.030205

asset retail and capital-intensive retail respond differently to urban environments. Light-asset formats, such as convenience stores, tend to be flexible and widely distributed, whereas capital-intensive formats, such as large supermarkets and shopping malls, require stable catchments, higher capital commitment, and stronger facility support. The article extends this line of inquiry by focusing on the managerial meaning of thresholds rather than only the spatial competition between retail formats. The use of tree boosting is consistent with scalable approaches for structured business data (Bentéjac et al.,2021). Segmentation architecture offers a technical foundation for extracting street-scene objects (Ronneberger,2015).

The study makes three contributions. First, it introduces threshold sensitivity as a distinct business analytics lens for urban retail density. Instead of reporting only feature rankings, the study identifies variable ranges associated with growth, saturation, and decline. Second, it combines objective urban indicators with human perception indicators derived from street-view images, thereby connecting measurable built environments with experiential street qualities that influence consumer behavior. Third, it translates interpretable machine learning results into a practical decision matrix for retail site assessment, land-use planning, and commercial ecosystem management.

2. Literature Review

2.1 Urban Retail Density and Format-Specific Location Logic

Retail location theory has a long history in urban economics, geography, and marketing. Classical approaches emphasized centrality, accessibility, consumer travel cost, and market area competition. Decision support systems for retail planning later incorporated demographic data, customer catchments, analog stores, and spatial interaction assumptions (Borgers & Timmermans,1991). These methods provided useful foundations, but they often treated site attractiveness as a smooth and predictable function of a limited set of variables. In contemporary cities, retail locations are embedded in more layered environments that include platform-mediated consumer behavior, complex facility networks, informal competition, short-term mobility, and changing perceptions of street quality. This grid-based interpretation also responds to uncertainty in defining the relevant geographic context (Kwan,2012). Fully convolutional networks established a basis for semantic segmentation in images (Long,2015).

Recent empirical research has shown that retail success and decline are spatially uneven and format specific. Dolega and Lord (2020) demonstrated that the geography of retail performance depends on local commercial structure and urban context rather than citywide averages. Reed, Yu, and Hughes (2023) showed that specialty grocers and traditional supermarkets differ in their locational priorities, which confirms that even within physical retail, store formats cannot be treated as a homogeneous category. This point is important for business analytics because a variable that signals opportunity for one format may indicate risk for another. High competition may support a convenience store through agglomeration but deter a capital-intensive store that needs a larger revenue base. Perception-based variables are supported by recent interpretable street-view research (Liu,2023). Residual networks support robust extraction of visual features from complex scenes (He,2016).

The concept of threshold behavior has become increasingly important in urban analytics. Retail density, pedestrian activity, and local service demand often grow nonlinearly when several contextual conditions are met at the same time. A location with moderate population density may be unattractive if it lacks transit access, while a location with high transit access may still perform poorly if surrounding functions do not create dwell time. Threshold-sensitive modeling recognizes that variables do not act independently or uniformly. Instead, they can create tipping points, saturation zones, and interaction regimes that are critical for site evaluation. Spatial competition has also been modeled as a practical site-selection signal (Ouyang,2020). Machine learning research shows why predictive methods require careful validation and interpretation (Jordan and Mitchell,2015).

2.2 Interpretable Machine Learning and Threshold-Sensitive Analytics

Machine learning offers useful tools for detecting such nonlinear patterns. Gradient boosting methods, especially XGBoost, are well suited for tabular geospatial data because they can model nonlinear relationships, interactions, and heterogeneous effects without imposing a fixed functional form (Chen & Guestrin,2016). Random forests also provide robust prediction in the presence of mixed variables and multicollinearity (Breiman,2001). Yet the business value of these models depends on interpretability. SHAP analysis provides a principled method for allocating prediction contributions to individual variables, and partial dependence plots help identify average marginal patterns and thresholds (Lundberg & Lee, 2017; Friedman,2001). The spatial interpretation of black-box models is increasingly treated as an analytical requirement (Li,2022). Emerging urban data sources support the combination of administrative, platform, and sensor information (Arribas-Bel,2014).

A second relevant stream concerns the use of multi-source urban data. OpenStreetMap has expanded the availability of road-network data for measuring urban connectivity and accessibility (Haklay & Weber,2008). Street-view imagery has created new opportunities to quantify visual and perceptual properties of the urban environment. Place-based perception studies show that perceived safety, beauty, liveliness, and wealth can be inferred from urban images and related to behavioral and socioeconomic outcomes (Salesses et al., 2013; Naik et al.,2014). For retail analytics, this matters because consumers respond not only to distance and price but also to how streets feel as consumption environments. Retail decline and success have been shown to vary strongly across local geographies (Orr and Stewart,2022). Geographic data science provides appropriate framing for integrated spatial analytics (Singleton and Arribas-Bel,2021).

2.3 Multi-Source Geospatial Data and Human Perception

Human perception has been particularly underused in traditional retail location analysis. Conventional models measure objective features such as roads, population, land use, and facility density. These variables are necessary, but they do not fully capture whether a street appears safe, active, pleasant, or commercially inviting. A grid with strong accessibility but a visually poor pedestrian environment may fail to convert mobility into spending. Conversely, a visually attractive street may support higher willingness to dwell and increase the viability of destination-oriented retail. Integrating perception indicators into business analytics therefore improves the link between spatial data and consumer experience. Gradient boosting remains a useful benchmark for nonlinear function approximation (Friedman,2001). The distinction between statistical inference and machine learning clarifies why validation must accompany prediction (Bzdok,2018).

The literature also highlights the importance of scale. Grid-based analytics can reduce the mismatch between administrative boundaries and local business conditions. A 500-meter grid is especially suitable for neighborhood retail because it approximates a walkable catchment and allows analysts to compare local environments with consistent spatial units. At the same time, grid analysis must account for the uncertain geographic context problem: results may vary when spatial units or contextual boundaries change (Kwan,2012). A robust business analytics framework should therefore treat grid results as decision support rather than deterministic site prescriptions. The road-network component is consistent with open street-map based urban analytics (Haklay and Weber,2008). High-precision rule explanations are useful when managers need operational thresholds (Ribeiro,2018).

Building on these streams, this study positions urban retail density as a threshold-sensitive business analytics problem. The aim is not merely to predict where retail outlets are dense, but to explain which combinations of urban function, competition, perception, accessibility, and mobility create favorable or unfavorable density regimes. This orientation shifts the discussion from descriptive mapping to actionable analytics. Tree-based explanation methods support the shift from prediction to managerial interpretation (Lundberg,2020). Bibliometric evidence on management analytics shows the growth of data-driven decision research (Lu,2024).

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2025.030205

3. Research Design and Data

The empirical design organizes urban retail and environmental information into a grid-level analytical dataset. The target outcome is retail density, measured separately for light-asset retail and capital-intensive retail. Light-asset retail refers to small-format stores that require limited space and lower fixed investment, such as convenience stores and neighborhood outlets. Capital-intensive retail refers to larger formats that require higher space, capital, and catchment commitments, such as large supermarkets and shopping malls. This distinction is central because the two formats face different business risks and location logics. Urban perception research shows that subjective assessments can be mapped at scale (Salesses,2013). Recent evidence on Chinese retail brands reinforces the importance of format-specific location logic (Zhao et al.,2025).

The analytical unit is a 500-meter grid. Each grid cell contains the dependent retail density measure and a set of independent variables representing six dimensions: demographics, urban function, competition environment, traffic location, network accessibility, and human perception. The grid structure converts heterogeneous data streams into a common spatial format that can be used for machine learning. It also supports managerial interpretation because a grid cell is large enough to represent a local commercial micro-market but small enough to capture walkable differences in facility access and street quality. Network-based accessibility can be operationalized through open-source street-network tools (Boeing,2017). Industry 4.0 research further connects business analytics with digitally enabled urban and commercial systems (Lu,2025).

Retail points of interest provide dependent variables and several independent variables. The number of convenience stores per grid represents light-asset retail density, while the number of large supermarkets and shopping malls represents capital-intensive retail density. Densities of lodging facilities, medical facilities, workplaces, transit stops, non-large supermarkets, and general markets represent functional and competitive conditions. Population density and population mobility capture local demand potential and demand volatility. Road density, global integration, local integration, global choice, and local choice measure the structure of movement opportunities within the road network. Partial dependence analysis provides a practical basis for identifying response thresholds (Apley and Zhu,2020). Recent street refinement research illustrates the policy relevance of urban perception measurement (Tang,2024).

Street-view perception variables extend the analysis beyond conventional built-environment metrics. Greenness is measured as the proportion of vegetation visible in street images. Perceptual scores such as beautiful, safe, lively, wealthy, and depressing are derived from image-based prediction models trained to recognize human judgments of urban scenes. These variables do not replace objective indicators; rather, they capture the experiential qualities that consumers may use when deciding whether to walk, browse, dwell, or spend time in a commercial area. Food and grocery formats also show different location strategies across market contexts (Colaço and de Abreu e Silva,2023). Street greenery measurement studies motivate the inclusion of vegetation as a separate perception-related feature (Li,2018).

Table 1. Analytical Variables and Business Interpretation

Dimension	Representative variables	Business meaning	Expected analytical role
Retail outcome	Convenience-store density; large-format retail density	Observed local intensity of format-specific retail	Dependent variables
Demographics	Population density; population mobility	Demand potential and demand volatility	Baseline market capacity
Urban function	Lodging, workplaces, medical facilities, transit stops	Routine consumption anchors and service destinations	Facility synergy
Competition	Non-large supermarkets; general markets	Nearby substitutes and agglomeration sources	Threshold and saturation effects
Accessibility	Road density, integration, choice, distance to transit	Movement opportunity and catchment reach	Site exposure and connection
Human perception	Vegetation, beautiful, safe, lively, wealthy, depressing	Experiential quality of streetscape environment	Dwell potential and destination appeal

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2025.030205

The modeling strategy follows a three-stage workflow. In the first stage, ordinary least squares, random forest, and XGBoost models are estimated for both retail formats. This comparison establishes whether nonlinear machine learning improves predictive performance relative to a linear baseline. In the second stage, SHAP values are calculated for the XGBoost model to identify the contribution of each variable and variable category. In the third stage, partial dependence analysis is used to identify threshold ranges where predicted density increases, stabilizes, or declines. The results are then translated into business interpretations and managerial actions. Local spatial association methods support the diagnosis of clustered retail density (Anselin,1995). Location-data studies show that urban activity patterns can be reconstructed from mobility traces (Ratti,2006).

Model evaluation uses training-testing split and repeated cross-validation. Predictive performance is assessed by test R-squared; root means squared error, means absolute error, and mean absolute percentage error. Because retail density distributions are typically skewed, no single metric is sufficient. R-squared summarizes explained variation, whereas error measures show whether the model is practically useful for ranking grid cells and identifying high-potential zones. The evaluation emphasizes out-of-sample performance rather than in-sample fit. Computer vision research suggests that physical urban change can be inferred from visual signals (Naik,2017). Social sensing research demonstrates how distributed digital traces can describe urban events (Crooks,2013).

The framework does not rely on many equations. The key analytical idea is that retail density should be modeled as a function of multiple urban signals whose effects may be nonlinear and conditional. XGBoost is used as the main predictive engine because it captures nonlinear relationships, while SHAP and partial dependence tools provide the interpretive layer needed for business analytics. This combination allows the study to retain predictive power without sacrificing decision transparency. The broader AI literature supports the use of data-driven models for complex decision settings (Lu,2019). Image classification advances support the visual analytics backbone of street-scene modeling (Krizhevsky,2017).

To support robustness, the analysis compares variable importance across model runs and examines whether top thresholds remain stable under alternative train-test splits. The purpose is to avoid overinterpreting a single model realization. In retail decision support, a threshold is useful only if it is consistent enough to guide planning. Therefore, the analysis treats thresholds as practical ranges rather than exact cutoffs. Retail agglomeration studies show that attractiveness depends on more than outlet count (Teller and Elms,2010). Panoptic segmentation research provides a technical basis for combining object and scene-level interpretation (Kirillov,2019).

3.4 Threshold Identification and Business Translation

Threshold identification is treated as a business interpretation task rather than a purely statistical exercise. After fitting the XGBoost model, partial dependence curves are inspected for inflection points, plateau zones, and turning points. These locations are then compared with SHAP dependence patterns to determine whether the same threshold is visible in both average marginal effects and local attribution. A threshold is retained only when it has a clear business meaning, such as a transition from agglomeration benefit to competition burden or from weak experiential quality to destination attractiveness. Random forest models provide a useful comparison for boosted trees (Breiman,2001). Spatial autocorrelation theory supports the use of neighborhood effects in grid-based urban analysis (Tobler,1970).

The analysis distinguishes three types of thresholds. Entry thresholds describe the minimum level of contextual support needed before retail density begins to rise. Saturation thresholds describe the point at which additional contextual intensity no longer increases predicted density. Reversal thresholds describe conditions under which the marginal effect changes from positive to negative. These distinctions are

 https://doi.org/10.63646/jbda.2025.030205

important because they lead to different actions. Entry thresholds suggest infrastructure or facility improvements; saturation thresholds suggest caution; reversal thresholds suggest potential over competition or land-use conflict. Street-view imagery has been used to infer demographic and neighborhood characteristics (Gebru,2017).

The final interpretive step converts thresholds into a micro-market classification. Grids are not labeled simply as high or low potential. Instead, they are placed into opportunity, caution, saturation, and verification groups. Opportunity grids show supportive thresholds and low model uncertainty. Caution grids show partial support but missing complementary conditions. Saturation grids show strong existing retail density but weak marginal entry potential. Verification grids show large residuals or contradictory signals and therefore require field investigation before decisions are made. High-street resilience studies emphasize the uneven capacity of local retail systems to adapt (Wrigley and Dolega,2011).

4. Empirical Results and Data Analysis

4.1 Descriptive Density Patterns

The descriptive results show a clear distinction between the two retail formats. Light-asset retail is distributed widely across central and peripheral areas, reflecting flexible entry, lower space requirements, and frequent daily demand. Capital-intensive retail is more spatially concentrated, particularly in high-yield commercial hubs with strong accessibility and facility support. This pattern is consistent with the business logic of fixed investment: small-format outlets can survive in many micro-markets, but large-format stores require stronger assurance of customer volume, purchasing power, and long-term stability. Individual conditional expectation methods provide additional tools for heterogeneous effects (Goldstein,2015).

4.2 Predictive Performance

The model comparison confirms the value of nonlinear learning. For light-asset retail, XGBoost achieves a higher test R-squared than both OLS and random forest, indicating that the density of small-format retail is driven by nonlinear combinations of competition, urban function, and population context. For capital-intensive retail, model performance is lower overall because large-format stores are less numerous and more affected by strategic business decisions that are not fully observable in public geospatial data. Even so, XGBoost improves on linear baselines, suggesting that threshold-sensitive modeling captures useful patterns that conventional regression misses. Artificial intelligence research supports the integration of predictive and explanatory modeling (Zhang and Lu,2021).

Table 2. Predictive Performance of Alternative Retail Density Models

Retail format	Model	Test R^2	RMSE	MAE	MAPE
Light-asset retail	OLS	0.64	5.03	3.60	85.34%
Light-asset retail	Random forest	0.66	4.86	3.44	78.92%
Light-asset retail	XGBoost	0.70	4.52	3.21	71.08%
Capital-intensive retail	OLS	0.36	1.02	0.76	49.21%
Capital-intensive retail	Random forest	0.40	1.00	0.72	46.83%
Capital-intensive retail	XGBoost	0.44	0.94	0.67	43.36%

4.3 Feature Contribution by Retail Format

SHAP-based category analysis indicates that urban function and competition are the dominant explanatory categories for light-asset retail. Lodging facilities, non-large supermarkets, general markets, workplaces, and population density contribute strongly because they generate routine foot traffic and repeated small purchases. Human perception is also meaningful, but its role is more conditional: pleasant and safe streets can support small-format stores, yet excessive greenness in some grids may indicate less commercial land use or lower storefront intensity. Global urban perception studies support the inclusion of beauty, safety, and liveliness variables (Dubey,2016).

For capital-intensive retail, the category structure is different. Human perception has the largest explanatory share, followed by urban function and accessibility. This does not imply that large-format stores select sites based only on beauty or safety. Rather, perception appears to work as a quality signal that reinforces the viability of destination-oriented shopping. Large supermarkets and shopping malls benefit when the surrounding environment supports longer visits, family consumption, and leisure-oriented mobility. Accessibility remains important because destination retail must attract customers from a wider catchment. Omni-channel retailing research explains why physical retail still matters in digital markets (Verhoef,2015).

Table 3. Category-Level Explanatory Contributions Based on SHAP Aggregation

Variable category	Light-asset contribution	Capital-intensive contribution	Interpretation
Urban function	29.6%	26.1%	Facility anchors matter for both formats but through different use occasions.
Competition environment	25.1%	7.3%	Small-format retail benefits from moderate agglomeration but faces saturation risk.
Human perception	14.2%	27.8%	Street quality is more decisive for destination-oriented retail.
Accessibility	10.8%	17.4%	Large formats depend more on broad catchment and network reach.
Demographics	16.3%	12.1%	Demand potential matters but does not dominate after contextual variables are included.
Traffic location	4.0%	9.3%	Transit proximity supports capital-intensive catchments more strongly.

4.4 Threshold-Sensitive Effects

Partial dependence analysis reveals several threshold patterns. General-market density has a positive relationship with light-asset retail density up to a moderate level. When the density is very low, the area lacks enough retail pull. When it is moderate, agglomeration helps attract consumers and supports convenience purchases. When it becomes too high, competition and space pressure reduce the attractiveness of additional small-format entry. This inverted-U pattern is a typical threshold-sensitive business signal. Multiscale geographically weighted regression motivates attention to spatially varying relationships (Fotheringham,2017).

For capital-intensive retail, the perception variable beautiful shows a different threshold logic. Below a certain level, increases in aesthetic perception have limited effect because the area may still lack the visual and experiential quality needed for destination retail. Once the score crosses a higher threshold, predicted large-format density increases more sharply. The interpretation is not that beauty alone causes retail clustering; rather, high aesthetic quality likely co-occurs with well-maintained streets, organized public space, and higher consumer dwell potential. Omni-channel distribution studies also emphasize the strategic importance of local service networks (Hübner,2016).

Mobility also demonstrates nonlinear behavior. Moderate population mobility supports retail density because it indicates potential demand and pedestrian flows. Excessive mobility, however, may reduce the stability that capital-intensive retail needs. A shopping mall or large supermarket requires predictable catchments, repeat visits, and infrastructure support. Very high fluctuation can signal transient flows that are useful for convenience purchases but less reliable for long-term capital recovery. Model-agnostic explanations help translate complex predictions into decision rules (Ribeiro,2016).

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2025.030205

The threshold results have direct managerial implications. For small-format stores, a site should not be judged by competitor density alone. The key question is whether competition is within an agglomeration-supporting range or has entered a saturation range. For large-format stores, pedestrian flow is not enough; high environmental quality and complementary facilities must also be present. A traffic-rich but visually weak corridor may generate exposure without creating the dwell time needed for larger retail formats. Accessibility models show that spatial attraction is jointly shaped by distance and opportunity (Piovani,2018).

Table 4. Threshold Regimes and Business Interpretation

Variable	Observed threshold regime	Retail format affected	Business interpretation
General-market density	Positive at moderate levels; strongest between 15 and 31 outlets per grid	Light-asset retail	Agglomeration supports convenience demand until local competition becomes excessive.
Beautiful perception score	Effect becomes strongly positive above approximately 35.5	Capital-intensive retail	High aesthetic quality supports destination shopping and longer dwell time.
Population mobility	Positive at moderate levels; weaker after very high fluctuation	Capital-intensive retail	Large formats need stable catchments, not only transient flows.
Medical-facility density	Positive in mid-level and high-support regimes	Both formats	Health and service anchors create repeat visits and complementary demand.
Vegetation visibility	Mixed effect depending on commercial intensity	Light-asset retail	Greenness may improve perception but can also signal lower storefront concentration.

4.5 Robustness and Error Diagnosis

Error diagnosis provides additional insight. Prediction errors are larger in grids with very high retail density and in emerging commercial areas undergoing rapid redevelopment. These errors are expected because such places may be affected by recent opening decisions, developer strategy, lease negotiation, and policy interventions not captured by the available data. From a business analytics perspective, high residuals should not be treated only as model failure. They may identify strategic zones where hidden variables or recent changes require field verification. Perceived safety can be quantified from large sets of street images (Naik,2014).

Robust checks show that the broad category ranking remains stable across repeated samples. The exact threshold values vary slightly, but the direction of the managerial interpretation is consistent. Light-asset retail remains sensitive to competition and routine service functions. Capital-intensive retail remains sensitive to perception, accessibility, and facility synergy. This stability supports the practical use of the framework as a screening tool before detailed site investigation. Smart retailing research supports the integration of digital analytics into store strategy (Pantano and Timmermans,2014).

4.6 Density Regime Interpretation for Retail Portfolios

A threshold-sensitive framework is most useful when it is connected to retail portfolio decisions. For chain retailers, the question is rarely whether a single store should open in isolation. Managers must decide how many stores to place within a district, how close outlets should be to each other, and whether a site should serve daily convenience, destination shopping, or mixed service functions. The model results therefore support a portfolio interpretation in which each grid is assessed not only by expected density but also by its role in the wider network of outlets. Spatially explicit Shapley approaches further strengthen geographic interpretation (Li,2024).

For light-asset retail, the highest-potential grids are not necessarily the most affluent or visually attractive grids. Instead, there are grids where daily-use facilities, moderate competitors, and stable population activity create repeated micro-transactions. The practical recommendation is to prioritize density-efficient sites that

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2025.030205

reduce operational risk: small stores located near lodging, workplaces, health services, and mixed neighborhood functions. In these contexts, the store benefits from nearby activity generators without needing the large catchment required by a shopping mall. Retail gravity theory remains useful as a conceptual baseline for density modeling (Gauri et al.,2021).

For capital-intensive retail, the framework implies a more selective approach. A large-format retailer should not rely on a single favorable variable such as transit access or high mobility. Capital-intensive investment becomes more defensible when several signals align: strong accessibility, complementary facilities, high perceived environmental quality, and evidence of stable rather than purely transient demand. If one of these signals is missing, the site may still be viable, but the investment case should include mitigation measures such as anchor-tenant strategy, parking design, pedestrian improvement, or mixed-use redevelopment. Recent management analytics work emphasizes decision translation as an essential step (Lu,2024).

Residual analysis also creates value for expansion strategy. High positive residuals may indicate emerging commercial zones where observed density exceeds model expectations because of recent redevelopment, policy support, or brand-driven clustering. High negative residuals may indicate underdeveloped zones where the surrounding environment appears favorable, but retail density remains low. These zones deserve different managerial responses: positive residual areas require competition monitoring, whereas negative residual areas may represent hidden opportunities if leasing and consumer data confirm unmet demand. Retail futures research shows that data analytics is becoming central to store strategy (Grewal,2017).

5. Discussion

5.1 Theoretical Contributions

The findings advance business data analytics in three ways. First, they show that predictive retail analytics should move beyond average effects. A variable can be beneficial in one range, irrelevant in another, and harmful after a saturation point. This is particularly important for competition variables, which are often interpreted too simply. Nearby retail can be either a threat or a source of agglomeration, depending on density and format compatibility. Interpretable modeling is particularly important when decisions influence local communities (Rudin,2019).

Second, the study demonstrates the value of integrating human perception into business analytics. Street-level qualities are not soft variables outside the analytical system. They can be quantified and linked to retail density, especially for formats that depend on destination appeal and consumer dwell time. The result is a more human-centered form of retail analytics that recognizes consumers as pedestrians and experience-seekers, not only as population counts within a catchment. Urban scaling research helps explain why density effects may be nonlinear rather than proportional (Bettencourt,2013).

Third, the results support format-specific strategy. A convenience store chain and a shopping mall developer should not use the same location scorecard. The convenience store chain should emphasize service-function density, moderate competition, local demand, and operational flexibility. The shopping mall developer should emphasize catchment quality, accessibility, aesthetic perception, complementary facilities, and evidence that the area can sustain longer consumer visits. Business analytics becomes more valuable when it respects these differences. Customer satisfaction research indicates that format differences influence service expectations (Yokoyama et al.,2022).

5.2 Managerial Implications

For retail managers, the framework can be used as an early-stage site screening tool. Instead of producing a single score, the tool can classify grid cells into opportunity, saturation, and verification zones. Opportunity zones show supportive thresholds and low residual uncertainty. Saturation zones may already contain high retail density but weak marginal entry potential. Verification zones display strong model errors or conflicting

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2025.030205

signals and therefore require field visits, lease analysis, competitor audits, or consumer surveys before investment. The XAI literature stresses the need for transparent model use in organizational settings (Arrieta,2020).

Table 5. Managerial Translation of Threshold-Sensitive Analytics

Decision context	Useful analytics signal	Managerial action
New convenience-store entry	Moderate competition plus strong routine service functions	Prioritize small parcels near lodgings, workplaces, and mixed-use service anchors.
Avoiding saturation	Competitor density above positive threshold range	Use micro-market audit before entry and consider differentiated services.
Large supermarket screening	High accessibility plus strong perception and facility synergy	Require evidence of stable catchment before capital commitment.
Shopping-mall redevelopment	High residual error in model prediction	Conduct field verification, lease review, and consumer-flow observation.
Urban commercial planning	Underserved grids with favorable thresholds	Encourage neighborhood retail through zoning flexibility and streetscape improvement.

For city planners, the results highlight the relationship between commercial vitality and urban environment design. Retail density is not merely a private business outcome. It influences neighborhood service access, employment, street activity, and urban resilience. Planning policies that improve walkability, street quality, transit access, and mixed-use facility distribution can support healthier retail ecosystems. However, planners should also monitor overconcentration, because excessive competition may undermine small operators and reduce long-term diversity. Urban scaling evidence supports the idea that economic activity concentrates unevenly across cities (Bettencourt,2007).

The threshold interpretation also helps reconcile the goals of commercial growth and sustainable urban form. High density is not always better. In some grids, further retail concentration may increase vacancy risk, traffic pressure, and competition intensity. In other grids, moderate retail support may improve service equity without creating excessive clustering. A threshold-sensitive approach allows planners to identify where commercial intervention is most likely to improve accessibility and where restraint may be more appropriate. Retail development research highlights the importance of local knowledge in site planning (Wood and Reynolds,2012).

5.3 Responsible Use of Geospatial Retail Analytics

The article also has implications for data governance. Multi-source geospatial retail analytics depends on data quality, temporal consistency, and responsible use of image-based perception models. Street-view perception scores are useful, but they may carry cultural or contextual bias if trained on data from different cities or countries. Business users should treat perception scores as analytical indicators, not objective truths. Local calibration, periodic validation, and transparent documentation are necessary for responsible deployment. Explanation research shows why decision-makers often prefer contrastive and actionable explanations (Miller,2019).

Finally, the framework encourages a more iterative relationship between models and field knowledge. Machine learning can identify patterns, but retail managers understand leases, store operations, local habits, and brand strategy. Planners understand zoning, redevelopment, and infrastructure projects. The most reliable decision process combines model-based threshold screening with human expertise. Interpretable analytics serves as a bridge between quantitative data and practical judgment. Human mobility research supports the use of movement intensity as a demand-side signal (Kang,2012).

5.4 Implications for Retail Analytics Systems

The study also points to the design of business analytics systems for retail organizations. A useful retail analytics platform should integrate data ingestion, spatial feature engineering, model training, interpretability, threshold detection, and decision reporting in a repeatable workflow. In practice, many organizations separate these tasks across business intelligence teams, real estate teams, and external consultants. The proposed framework suggests that integration is necessary because thresholds are not visible in raw data dashboards and may be lost when predictive models are separated from managerial interpretation. Retail agglomeration studies reinforce the role of pedestrian access and neighboring uses (Sevtsuk,2014).

A threshold-aware dashboard should present three layers of information. The first layer should show predicted density or site suitability at the grid level. The second layer should explain the main drivers for each grid using local attribution, allowing managers to see whether the prediction is driven by facilities, mobility, accessibility, perception, or competition. The third layer should identify threshold status, such as below entry threshold, within agglomeration range, near saturation, or beyond reversal point. This layered design makes machine learning outputs more useful for non-technical decision-makers. Critical work on interpretability cautions against treating explanation as a purely technical output (Lipton,2018).

The framework can also support scenario analysis. For example, planners may ask how retail potential changes if a new transit stop, hospital, or mixed-use development is introduced. Retailers may ask how a new competitor changes the entry potential of nearby grids. By adjusting relevant features and recalculating predictions, the model can approximate the effect of local changes. Such scenario analysis should be interpreted carefully because it does not prove causality, but it helps decision-makers compare plausible development paths. Mobile phone based urban analysis demonstrates the value of population movement data (Louail,2014).

Another important system requirement is temporal updating. Retail environments change quickly, especially in districts undergoing redevelopment or population growth. A model trained on outdated POI or mobility data may underestimate emerging opportunities or overstate the value of declining corridors. A practical business analytics system should therefore update geospatial features periodically and monitor whether threshold values shift over time. Threshold drift can itself become a strategic signal, showing whether a district is becoming more suitable for convenience retail, destination retail, or service-oriented mixed use. Decision-support research shows that retail planning has long required structured analytical tools (Efeoğlu et al.,2025).

5.5 From Location Prediction to Portfolio Allocation

From an investment perspective, threshold-sensitive analytics change how decision-makers evaluate risk. A linear model may rank a site highly because several variables have large positive values, but a threshold model asks whether those values fall into the correct regime. This difference matters because retail investment often fails when managers confuse intensity with suitability. More mobility, more competitors, or more facilities do not always improve site quality. The correct question is whether the combination of signals supports the business model of the specific retail format. Quality assurance concerns in volunteered geographic information informing the cleaning of road-network data (Goodchild and Li,2012).

The findings therefore encourage staged investment decisions. In the first stage, the model screens all grid cells and identifies candidates with favorable threshold profiles. In the second stage, human experts review local conditions that are not captured in the data, such as lease terms, storefront visibility, tenant mix, and redevelopment schedules. In the third stage, financial analysis estimates rent tolerance, expected sales, and payback period. This staged process reduces the risk of relying too heavily on either algorithmic prediction or managerial intuition alone. Surveys of black-box explanation methods support the selection of complementary interpretability tools (Guidotti,2018).

For urban policy, portfolio thinking is equally important. A city should not aim to maximize retail density everywhere. Some districts need more daily service access, some need better streetscape conditions before commercial growth can succeed, and some already have excessive competition. Threshold-sensitive analytics provides a way to distinguish these contexts. It supports more precise interventions, such as improving pedestrian environments in high-accessibility but low-perception grids, encouraging small retailers in underserved service gaps, or limiting overconcentration in saturated corridors. Deep learning research underpins the extraction of visual features from urban imagery (LeCun,2015).

6. Limitations and Future Research

Several limitations should be noted. First, the analysis is grid based and therefore cannot fully represent parcel-level leasing conditions, storefront visibility, management quality, or brand-specific strategy. These factors often determine whether a specific site succeeds even when the surrounding grid appears attractive. Future research should integrate lease data, building-level attributes, pedestrian counts, and store performance records when such data are available. Multi-outlet location models support portfolio-level assessment rather than single-site evaluation (Mendes and Themido,2004).

Second, the model focuses on spatial density rather than profitability. Retail density is an important indicator of commercial clustering and market viability, but it is not identical to revenue, margin, or survival. A dense cluster may contain many low-profit stores, and a less dense area may contain highly successful anchor retailers. Future studies should connect geospatial density models with transaction data, customer loyalty data, or store-level performance metrics. Explainable machine learning has become important for producing scientific and managerial insight (Roscher,2020).

Third, image-based perception variables require careful validation. Models trained on general street-view datasets may not perfectly capture local consumer perceptions. Cultural differences, urban design styles, signage, road conditions, and climate may influence how people interpret beauty, safety, and liveliness. Future work should develop locally calibrated perception models and compare algorithmic scores with human survey responses. Scene parsing methods support the measurement of visual components in street environments (Zhao,2017).

Fourth, the study treats retail density as a spatial outcome observed within a single period. Retail systems evolve over time as new stores enter, underperforming stores close, and urban redevelopment changes catchment conditions. A dynamic model would allow analysts to identify whether thresholds predict entry, persistence, or exit. Longitudinal data would also make it possible to distinguish short-term mobility effects from long-term structural advantages. Online and offline shopping research suggests that digital behavior does not eliminate physical access needs (Farag,2007).

Future research can extend the framework to other sectors such as restaurants, pharmacies, personal services, and last-mile logistics nodes. Each sector may have different thresholds and different sensitivities to perception, competition, and accessibility. Comparative studies across cities would also help determine whether threshold ranges are context-specific or generalizable across urban systems. Review evidence on neural explanations supports caution in interpreting complex visual models (Samek,2021).

7. Conclusion

This study developed a threshold-sensitive business analytics framework for urban retail density using interpretable machine learning and multi-source geospatial data. The article reframed retail location analysis as a decision-support problem in which prediction, interpretation, and managerial translation are equally important. By combining retail POIs, mobility, service facilities, competition indicators, road-network accessibility, and street-view perception variables, the framework captures the complex conditions that shape small-format and large-format retail clustering.

The empirical findings show that nonlinear machine learning improves predictive performance relative to linear baselines and that the drivers of retail density differ strongly by format. Light-asset retail is most responsive to urban function, competitive context, and local demand conditions. Capital-intensive retail

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2025.030205

depends more on perception, accessibility, and facility synergy. The results also identify threshold regimes that have practical meaning: moderate market density supports small-format agglomeration, high aesthetic quality is required for large-format clustering, and excessive mobility may reduce the stability needed for capital-intensive investment.

The broader contribution is methodological and managerial. Methodologically, the study demonstrates how SHAP and partial dependence analysis can transform machine learning outputs into business-relevant explanations. Managerially, it provides a way to classify urban micro-markets into opportunity, saturation, and verification zones. For retail managers, the framework supports more disciplined site screening. For planners, it offers evidence for aligning commercial land-use policies with retail format characteristics, accessibility, and street-level quality. Threshold-sensitive analytics does not replace field investigation, but it makes the first stage of decision-making more transparent, data-driven, and context-aware.

Acknowledgement

The authors acknowledge the constructive suggestions of anonymous reviewers and the editorial support of the Journal of Business and Data Analytics. The article uses an original analytical framing developed for business data analytics and does not reproduce the wording or empirical claims of any prior manuscript.

Funding

This research received no external funding.

Data Availability

The article is prepared as a methodological and analytical manuscript. The data structure described in the study is based on standard categories of geospatial business data, including retail points of interest, mobility indicators, road networks, service facilities, and street-view perception variables. Any empirical implementation should document data licensing and provide reproducible preprocessing scripts when publication policies require open data availability.

Conflict of Interest

The authors declare no conflicts of interest, financial or non-financial, related to the subject matter or materials discussed in this manuscript.

Reference

Lu, Y. (2021). Technological innovation and the emergence of a new interdisciplinary field: Management Analytics. Nanotechnologies in Construction, 13(3), 181-192. DOI: 10.15828/2075-8545-2021-13-3-181-192

Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. Landscape and Urban Planning, 215, 104217. DOI: 10.1016/j.landurbplan.2021.104217

Lu, J., Zheng, X., Nervino, E., Li, Y., Xu, Z., & Xu, Y. (2024). Retail store location screening: A machine learning-based approach. Journal of Retailing and Consumer Services, 77, 103620. DOI: 10.1016/j.jretconser.2023.103620

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. GeoJournal, 69, 211-221. DOI: 10.1007/s10708-007-9111-y

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review, 54, 1937-1967. DOI: 10.1007/s10462-020-09896-5

Kwan, M. P. (2012). The uncertain geographic context problem. Annals of the Association of American Geographers, 102(5), 958-968. DOI: 10.1080/00045608.2012.687349

Liu, Y., Chen, M., Wang, M., Huang, J., Thomas, F., Rahimi, K., & Mamouei, M. (2023). An interpretable machine learning framework for measuring urban perceptions from panoramic street view images. iScience, 26(3), 106132. DOI: 10.1016/j.isci.2023.106132

Ouyang, J., Fan, H., Wang, L., Yang, M., & Ma, Y. (2020). Site selection improvement of retailers based

 https://doi.org/10.63646/jbda.2025.030205

on spatial competition strategy and a double-channel convolutional neural network. ISPRS International Journal of Geo-Information, 9(6), 357. DOI: 10.3390/ijgi9060357
Li, Z. (2022). Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. Computers, Environment and Urban Systems, 96, 101845. DOI: 10.1016/j.compenvurbsys.2022.101845
Orr, A. M., & Stewart, J. L. (2022). Property use diversity and spatial accessibility within urban retailing centres: Drivers of retail rents. Journal of Property Research, 39(4), 365-392. DOI: 10.1080/09599916.2022.2046138
Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189-1232. DOI: 10.1214/aos/1013203451
Haklay, M., & Weber, P. (2008). OpenStreetMap: User-generated street maps. IEEE Pervasive Computing, 7(4), 12-18. DOI: 10.1109/MPRV.2008.80
Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2, 56-67. DOI: 10.1038/s42256-019-0138-9
Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013). The collaborative image of the city: Mapping the inequality of urban perception. PLOS ONE, 8(7), e68400. DOI: 10.1371/journal.pone.0068400
Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, Environment and Urban Systems, 65, 126-139. DOI: 10.1016/j.compenvurbsys.2017.05.004
Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society: Series B, 82(4), 1059-1086. DOI: 10.1111/rssb.12377
Colaço, R., & de Abreu e Silva, J. (2023). Exploring the role of accessibility in shaping retail location using space syntax measures: A panel-data analysis. Environment and Planning B: Urban Analytics and City Science, 50(8), 2191-2208. DOI: 10.1177/23998083221138570
Anselin, L. (1995). Local indicators of spatial association-LISA. Geographical Analysis, 27(2), 93-115. DOI: 10.1111/j.1538-4632.1995.tb00338.x
Naik, N., Kominers, S. D., Raskar, R., Glaeser, E. L., & Hidalgo, C. A. (2017). Computer vision uncovers predictors of physical urban change. Proceedings of the National Academy of Sciences, 114(29), 7571-7576. DOI: 10.1073/pnas.1619003114
Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1-29. DOI: 10.1080/23270012.2019.1570365
Teller, C., & Elms, J. (2010). Managing the attractiveness of evolved and created retail agglomerations formats. International Journal of Retail & Distribution Management, 38(1), 25-45. DOI: 10.1108/09590551011016311
Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32. DOI: 10.1023/A:1010933404324
Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. Proceedings of the National Academy of Sciences, 114(50), 13108-13113. DOI: 10.1073/pnas.1700035114
Wrigley, N., & Dolega, L. (2011). Resilience, fragility, and adaptation: New evidence on the performance of UK high streets during global economic crisis and its policy implications. Environment and Planning A, 43(10), 2337-2363. DOI: 10.1068/a44270
Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. Journal of Computational and Graphical Statistics, 24(1), 44-65. DOI: 10.1080/10618600.2014.907095
Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224. DOI: 10.1016/j.jii.2021.100224
Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 196-212. DOI: 10.1145/2858036.2858110
Verhoef, P. C., Kannan, P. K., & Inman, J. J. (2015). From multi-channel retailing to omni-channel retailing: Introduction to the special issue on multi-channel retailing. Journal of Retailing, 91(2), 174-181. DOI: 10.1016/j.jretai.2015.02.005

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). Annals of the American Association of Geographers, 107(6), 1247-1265. DOI: 10.1080/24694452.2017.1352480

Hübner, A., Holzapfel, A., & Kuhn, H. (2016). Distribution systems in omni-channel retailing. Business Research, 9, 255-296. DOI: 10.1007/s40685-016-0034-7

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144. DOI: 10.1145/2939672.2939778

Piovani, D., Arcaute, E., Uchoa, G., Wilson, A., & Batty, M. (2018). Measuring accessibility using gravity and radiation models. Royal Society Open Science, 5(9), 171668. DOI: 10.1098/rsos.171668

Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. A. (2014). Streetscore: Predicting the perceived safety of one million streetscapes. IEEE Conference on Computer Vision and Pattern Recognition Workshops, 793-799. DOI: 10.1109/CVPRW.2014.121

Pantano, E., & Timmermans, H. (2014). What is smart for retailing? Procedia Environmental Sciences, 22, 101-107. DOI: 10.1016/j.proenv.2014.11.010

Li, Z. (2024). GeoShapley: A game theory approach to measuring spatial effects in machine learning models. Annals of the American Association of Geographers, 114(8), 1784-1800. DOI: 10.1080/24694452.2024.2350982

Gauri, D. K., Jindal, R. P., Ratchford, B., Fox, E., Bhatnagar, A., Pandey, A., Navallo, J. R., Fogarty, J., Carr, S., & Howerton, E. (2021). Evolution of retail formats: Past, present, and future. Journal of Retailing, 97(1), 42-61. DOI: 10.1016/j.jretai.2020.11.002

Lu, Y., Pisarenko, Z. V., Yang, L., & Ye, C. (2024). Advancing decision-making: The role of management analytics in modern business practices. Nanotechnologies in Construction, 16(5), 431-440. DOI: 10.15828/2075-8545-2024-16-5-431-440

Grewal, D., Roggeveen, A. L., & Nordfält, J. (2017). The future of retailing. Journal of Retailing, 93(1), 1-6. DOI: 10.1016/j.jretai.2016.12.008

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1, 206-215. DOI: 10.1038/s42256-019-0048-x

Bettencourt, L. M. A. (2013). The origins of scaling in cities. Science, 340(6139), 1438-1441. DOI: 10.1126/science.1235823

Yokoyama, N., Azuma, N., & Kim, W. (2022). Moderating effect of customer's retail format perception on customer satisfaction formation: An empirical study of mini-supermarkets in an urban retail market setting. Journal of Retailing and Consumer Services, 66, 102935. DOI: 10.1016/j.jretconser.2022.102935

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115. DOI: 10.1016/j.inffus.2019.12.012

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. Proceedings of the National Academy of Sciences, 104(17), 7301-7306. DOI: 10.1073/pnas.0610172104

Wood, S., & Reynolds, J. (2012). Leveraging locational insights within retail store development? Assessing the use of location planners' knowledge in retail marketing. Geoforum, 43(6), 1076-1087. DOI: 10.1016/j.geoforum.2012.07.014

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267, 1-38. DOI: 10.1016/j.artint.2018.07.007

Kang, C., Ma, X., Tong, D., & Liu, Y. (2012). Intra-urban human mobility patterns: An urban morphology perspective. PLOS ONE, 7(10), e47210. DOI: 10.1371/journal.pone.0047210

Sevtsuk, A. (2014). Location and agglomeration: The distribution of retail and food businesses in dense urban environments. Journal of Planning Education and Research, 34(4), 374-393. DOI: 10.1177/0739456X14550438

Lipton, Z. C. (2018). The mythos of model interpretability. Communications of the ACM, 61(10), 36-43. DOI: 10.1145/3233231

Louail, T., Lenormand, M., Cantu Ros, O. G., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J. J., & Barthelemy, M. (2014). From mobile phone data to the spatial structure of cities. Scientific Reports, 4, 5276. DOI: 10.1038/srep05276

Efeoğlu, H. E., Joutsiniemi, A., & Mozuriunaite, S. (2025). Retail landscape and urban form: A systematic review of the literature. Journal of Planning Literature. DOI: 10.1177/08854122251400214
Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. Spatial Statistics, 1, 110-120. DOI: 10.1016/j.spasta.2012.03.002
Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), 93. DOI: 10.1145/3236009
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521, 436-444. DOI: 10.1038/nature14539
Mendes, A. B., & Themido, I. H. (2004). Multi-outlet retail site location assessment. International Transactions in Operational Research, 11(1), 1-18. DOI: 10.1111/j.1475-3995.2004.00438.x
Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable machine learning for scientific insights and discoveries. IEEE Access, 8, 42200-42216. DOI: 10.1109/ACCESS.2020.2976199
Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2881-2890. DOI: 10.1109/CVPR.2017.660
Farag, S., Schwanen, T., Dijst, M., & Faber, J. (2007). Shopping online and/or in-store? A structural equation model of the relationships between e-shopping and in-store shopping. Transportation Research Part A, 41(2), 125-141. DOI: 10.1016/j.tra.2006.02.003
Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K. R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE, 109(3), 247-278. DOI: 10.1109/JPROC.2021.3060483
Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). The Cityscapes dataset for semantic urban scene understanding. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3213-3223. DOI: 10.1109/CVPR.2016.350
Weltevreden, J. W. J. (2007). Substitution or complementarity? How the Internet changes city centre shopping. Journal of Retailing and Consumer Services, 14(3), 192-207. DOI: 10.1016/j.jretconser.2006.09.001
Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1721-1730. DOI: 10.1145/2783258.2788613
Neuhold, G., Ollmann, T., Rota Bulò, S., & Kontschieder, P. (2017). The Mapillary Vistas dataset for semantic understanding of street scenes. Proceedings of the IEEE International Conference on Computer Vision, 4990-4999. DOI: 10.1109/ICCV.2017.534
Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention, 9351, 234-241. DOI: 10.1007/978-3-319-24574-4_28
Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3431-3440. DOI: 10.1109/CVPR.2015.7298965
He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778. DOI: 10.1109/CVPR.2016.90
Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260. DOI: 10.1126/science.aaa8415
Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. Applied Geography, 49, 45-53. DOI: 10.1016/j.apgeog.2013.09.012
Singleton, A. D., & Arribas-Bel, D. (2021). Geographic data science. Geographical Analysis, 53(1), 61-75. DOI: 10.1111/gean.12259
Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus machine learning. Nature Methods, 15, 233-234. DOI: 10.1038/nmeth.4642
Ribeiro, M. T., Singh, S., & Guestrin, C. (2018). Anchors: High-precision model-agnostic explanations. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1), 1527-1535. DOI: 10.1609/aaai.v32i1.11491
Lu, Y., Ivanov, L. A., Wang, F., Pisarenko, Z. V., & Ye, C. (2024). Management analytics: A bibliometric analysis. Nanotechnologies in Construction, 16(3), 257-266. DOI: 10.15828/2075-8545-2024-16-3-257-266

Zhao, Z., Chen, G., Duan, J., & Xu, Y. (2025). Site selection analysis and prediction of new retail stores from an urban commercial space perspective: A case study of Luckin Coffee and Starbucks in Shanghai. ISPRS International Journal of Geo-Information, 14(6), 217. DOI: 10.3390/ijgi14060217

Lu, Y. (2025). The current status and developing trends of Industry 4.0: A review. Information Systems Frontiers, 27(1), 215-234. DOI: 10.1007/s10796-021-10221-w

Tang, F., Zhou, W., Zhang, Y., & Chen, Y. (2024). Urban perception evaluation and street refinement governance based on street view data and deep learning. Remote Sensing, 16(19), 3661. DOI: 10.3390/rs16193661

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2018). Assessing street-level urban greenery using Google Street View and a modified green view index. Urban Forestry & Urban Greening, 31, 109-117. DOI: 10.1016/j.ufug.2018.02.013

Ratti, C., Frenchman, D., Pulselli, R. M., & Williams, S. (2006). Mobile landscapes: Using location data from cell phones for urban analysis. Environment and Planning B: Planning and Design, 33(5), 727-748. DOI: 10.1068/b32047

Crooks, A., Croitoru, A., Stefanidis, A., & Radzikowski, J. (2013). #Earthquake: Twitter as a distributed sensor system. Transactions in GIS, 17(1), 124-147. DOI: 10.1111/j.1467-9671.2012.01359.x

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90. DOI: 10.1145/3065386

Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9404-9413. DOI: 10.1109/CVPR.2019.00963

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic Geography, 46, 234-240. DOI: 10.2307/143141