DATAMIND | Journal of Data-Driven AI and Computational Discovery

Vol. 3, No. 4, pp.: 50-70, 2025

A Geospatial Retail Intelligence Database for Urban Commercial Planning: Integrating POIs, Street Views, Mobility Heat Maps, and Road Networks

Haochen Lin[1], **Meiyu Zhang**[2,*], **Yifan Chen**[3]

[1] Department of Information Management, Guangzhou University, Guangzhou 510006, China

[2] School of Management, Guangdong University of Finance and Economics, Guangzhou 510320, China

[3] Department of Urban Informatics, Dongguan University of Technology, Dongguan 523808, China

* *meiyu.zhang@gdufe.edu.cn*

Article Information	
Received	17 July 2025
Accepted	19 November 2025
DOI	https://doi.org/10.63646/datamind.2025.030404

Abstract

Urban commercial planning increasingly depends on data infrastructures that translate heterogeneous spatial signals into actionable knowledge. Existing retail-location studies often rely on one data family, such as points of interest or transport accessibility, and therefore underrepresent the combined influence of human mobility, streetscape perception, facility synergy, and network structure. This article develops a geospatial retail intelligence database for urban commercial planning by integrating retail and facility POIs, street-view perception features, mobility heat maps, and road-network indicators into a governed spatial feature store. The proposed database is designed around 500-meter planning grids, multi-source metadata, repeatable quality-control rules, and interpretable analytical outputs for retail density assessment. Using an illustrative Shenzhen-style urban dataset, the study demonstrates how the database supports density benchmarking, threshold-sensitive feature analysis, and scenario-oriented planning for light-asset and capital-intensive retail formats. The results show that a unified geospatial database improves planning interpretability by linking facility proximity, accessibility, population dynamics, and perceptual streetscape conditions. The article contributes a database-centered framework for transforming scattered urban data into a reusable commercial intelligence asset for planners, retailers, and data-driven AI applications.

Keywords: ***Geospatial database; retail intelligence; urban commercial planning; POI analytics; street-view perception; mobility heat map; road network; data-driven AI***

1. Introduction

Urban retail systems are changing from location-driven business ecologies into data-observable spatial systems. A convenience store, supermarket, or shopping mall is no longer assessed only by visible foot traffic or rent. Its potential is increasingly interpreted through high-resolution POI distributions, mobile population signals, street-view imagery, transit access, and network centrality. For urban commercial planning, this transition creates both opportunity and fragmentation. The opportunity is that planners can observe the city at a finer spatial scale than traditional land-use maps permitted. The fragmentation is that retail-relevant data are usually stored in incompatible formats, collected by different platforms, and updated at different temporal frequencies. A planning department may have zoning layers, a retailer may have transaction records, a mapping company may hold POIs, and a mobility provider may hold heat-map data, yet these assets rarely converge into a stable analytical database. This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Lundberg et al.,2020). The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Hastie et al.,2009).

The research direction motivating this article is the application of interpretable machine learning to traditional retail location competition. The uploaded manuscript on Shenzhen retail density demonstrates that light-asset and capital-intensive formats respond differently to urban function, competition, accessibility, and human perception. That study treats multi-source geospatial data as inputs to predictive and explanatory models. This article shifts the emphasis from model output to database design. It asks how a reusable geospatial retail intelligence database should be structured so that future models, dashboards, audits, and planning scenarios can be generated from the same governed data foundation. The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Janowicz et al.,2020). This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Liu et al.,2015).

The distinction matters because a single model, even when accurate, does not provide institutional memory. Urban commercial planning requires repeatable evidence across years, districts, retail categories, and policy scenarios. A database-centered approach makes those comparisons possible. It separates data acquisition, cleaning, spatial harmonization, feature engineering, model training, and decision reporting into accountable layers. This separation allows planners to inspect why a grid cell is classified as under-served, why a retail corridor is judged saturated, or why a capital-intensive retail site requires stronger accessibility and streetscape quality than a convenience-store cluster. This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Biljecki et al.,2021). The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (Burkart et al.,2021).

A geospatial retail intelligence database should not be a passive repository. It should act as a planning-oriented feature store that preserves raw source lineage while producing standardized grid-level indicators. Each grid cell should become a spatial business record containing retail density, facility mix, human mobility, road accessibility, and perception scores. Such a record can serve descriptive benchmarking, predictive modeling, threshold discovery, and policy simulation. For example, a district government may compare whether a new neighborhood has enough daily-service retail relative to mobility intensity. A retail chain may evaluate where facility synergy and perceived vitality jointly support expansion. A data scientist may train an interpretable model without rebuilding the entire data pipeline. The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (Chen et al.,2016). This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (Porta et al.,2006).

This article therefore develops a database design and analytical demonstration for geospatial retail intelligence. It contributes in three ways. First, it proposes a normalized spatial data architecture that integrates POIs, street views, mobility heat maps, and road networks into a grid-level database. Second, it specifies quality-control procedures for completeness, temporal consistency, spatial joining, and feature validity. Third, it demonstrates how database outputs can support threshold-sensitive retail-density analysis without treating the model as a black box. The remainder of the article reviews related literature, presents the database methodology, reports analytical findings, discusses planning implications, and concludes with limitations and future research directions.

2. Literature Review and Theoretical Background

Retail location theory has long emphasized accessibility, market size, competition, and agglomeration. Classical spatial-interaction models explain why large stores gravitate toward high-demand centers while smaller daily-service outlets diffuse across neighborhoods. Recent data-rich studies extend this logic by showing that urban retail performance depends not only on population and transport but also on facility mix, platform visibility, and street-level environmental quality. The move from static retail geography to computational retail analytics is particularly important for fast-growing cities where land-use change, transit expansion, and consumer behavior evolve more quickly than traditional survey cycles. This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (Apley et al.,2020). This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Cheng et al.,2021).

Multi-source geospatial data have become central to this computational turn. POI datasets provide fine-grained representations of retail, employment, education, health care, accommodation, and transport facilities. Mobility heat maps provide dynamic indicators of population intensity and temporal fluctuation. Road networks and space-syntax indicators provide measures of connectivity and movement potential. Street-view images provide observable information about greenery, enclosure, safety, beauty, and vitality that cannot be captured by POIs alone. Together, these data families make it possible to measure commercial environments as interacting systems rather than isolated site attributes. This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Scutari,2010). The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Longley et al.,2016).

The database challenge is that each data family has a different spatial logic. POIs are point objects with category labels and uncertain update histories. Street-view images are directional visual samples whose perception features depend on sampling density and model inference quality. Mobility heat maps are raster-like or aggregate signals that may change by hour or day. Road networks are topological graphs with nodes, edges, and connectivity measures. Without a common indexing unit, these data cannot be compared reliably. Planning grids, such as 500-meter cells, provide a practical spatial unit because they approximate neighborhood walking distance while remaining large enough to stabilize density estimates. The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Zhang et al.,2018). This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Fotheringham et al.,2015).

The feature-store perspective has emerged in machine learning engineering as a way to prevent inconsistent feature construction across projects. In urban analytics, the same idea can support transparency. Instead of generating a new set of features for each paper or planning report, a city can maintain a governed feature layer with documented variable definitions, source dates, transformation rules, and missing-data flags. This is especially relevant for explainable AI because explanations are only meaningful when the underlying features are valid and comparable. A SHAP value for lodging density, for instance, is only interpretable if lodging POIs were collected consistently and aggregated using known spatial rules. This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Lu,2025). The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (Krizhevsky et al.,2017).

Urban perception research further strengthens the need for integrated databases. Street-view models trained on visual preference or semantic-segmentation data can estimate greenery, beauty, liveliness, safety, and perceived wealth. These variables often capture aspects of commercial attractiveness that conventional socioeconomic data miss. However, perception features can also be unstable if image capture dates, view angles, or model training domains differ. Database governance must therefore include metadata for image year, sampling method, perceptual model, and confidence level. Without these fields, perception scores risk becoming attractive but unaccountable numbers. The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (Ribeiro et al.,2016). This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (Zhou et al.,2019).

The theoretical position of this article is that retail intelligence should be treated as an urban data infrastructure problem before it is treated as a prediction problem. A well-designed database enables multiple models and planning uses, while a poorly documented database can make even sophisticated AI misleading. This position aligns with data-centric AI, which argues that performance and trust depend on improving data quality, labeling, lineage, and governance rather than only changing algorithms. In the retail planning context, data-centric AI means building an auditable spatial database in which each retail-density prediction is traceable to observable facility, mobility, perception, and network evidence. This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (Naik et al.,2017). This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Belle et al.,2021).

Table 1. Core Data Families and Planning Functions in the Retail Intelligence Database

Data family	Primary records	Planning function	Main quality concern
Retail and facility POIs	Store points, services, transit nodes	Density, competition, facility synergy	Category drift and duplicate records
Street-view images	Image samples and inferred attributes	Perception, greenery, visual quality	Sampling bias and model transferability
Mobility heat maps	Population intensity by time period	Demand rhythm and temporal stability	Temporal aggregation and privacy protection
Road networks	Edges, nodes, integration and choice indices	Accessibility and spatial structure	Topology breaks and inconsistent road classes
Planning grid	500-meter spatial cells	Common spatial index for all features	Boundary effects and low-count instability

Table 1 summarizes the database logic used in this article. The central design principle is not to force all source data into one raw format, but to keep source-specific records while producing a standardized planning grid that stores comparable analytical features. This allows each data family to preserve its original meaning and enables planners to evaluate the uncertainty attached to each feature.

3. Methodology and Database Design

The proposed database is organized as a layered spatial feature store. The source layer preserves raw or minimally processed POIs, image metadata, mobility observations, and road-network geometries. The governance layer stores source provider, acquisition date, licensing notes, coordinate system, refresh cycle, and quality flags. The harmonization layer converts records into a shared coordinate system and assigns them to a 500-meter grid. The feature layer stores normalized variables such as retail density, lodging density, medical-facility density, transit-stop density, population mobility, road density, global integration, local choice, vegetation, perceived beauty, perceived safety, and perceived liveliness. The analytical layer supports benchmarking, interpretable machine learning, threshold analysis, and planning dashboards. This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Boeing,2017). The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Darwiche,2009).

Figure 1 presents the architecture of the proposed database. It avoids an arrow-heavy process diagram and instead uses a layered layout to emphasize the database as an organized knowledge infrastructure. The architecture supports both descriptive and predictive tasks. A planner interested in equity can use the feature layer to compare underserved neighborhoods. A data scientist interested in model interpretation can use the analytical services layer to produce SHAP, PDP, or Bayesian-network outputs. A retailer can use the same records to screen candidate locations without rebuilding source-specific pipelines.

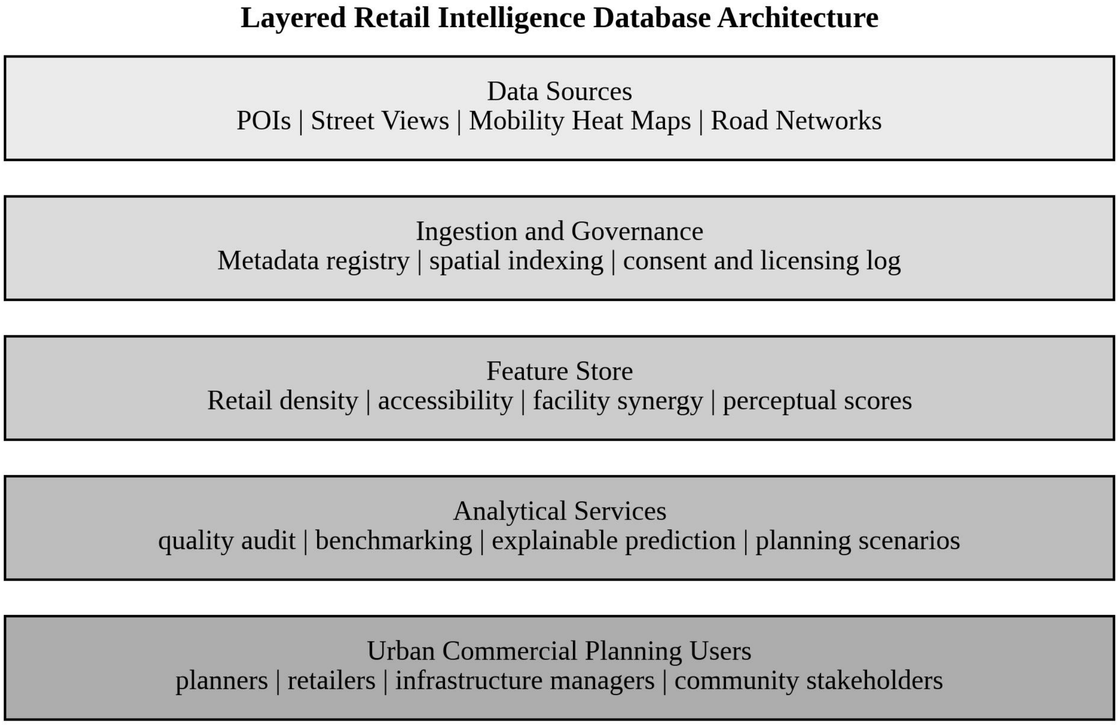

Figure 1. Layered retail intelligence database architecture for urban commercial planning.

The database uses grid cells as the central analytical records. A 500-meter grid is appropriate for neighborhood-scale retail analysis because it approximates a walkable catchment in dense urban districts while limiting random variation from individual addresses. For each grid, the database records counts of light-asset retailers, capital-intensive retailers, and surrounding facilities. It also records mobility statistics such as mean population intensity and day-to-day variation. Road-network features are computed from the network graph and attached to the grid by spatial overlay. Street-view perception features are aggregated from image samples located within or near the grid. The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Greenwell,2017). This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Kirillov et al.,2019).

A key design choice is to maintain both raw counts and standardized features. Raw counts are useful for direct planning interpretation: a grid contains a certain number of convenience stores, medical facilities, or transit stops. Standardized features are useful for modeling because they place variables with different units on comparable scales. The database therefore stores both the original variable and a z-score or min-max scaled version. Missing values are not silently imputed. Instead, the database records a missingness flag and an imputation method when imputation is used, allowing models and reports to distinguish observed absence from data absence. This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Adadi et al.,2018). The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (Korb et al.,2010).

Quality control is implemented through four groups of rules. Spatial rules test whether points fall inside the study boundary, whether road segments are connected, and whether images are correctly georeferenced. Semantic rules test whether POI categories are mapped consistently to retail and facility classes. Temporal rules test whether sources are close enough in acquisition time to support joint analysis. Statistical rules identify outliers, low-coverage grids, duplicated records, and extreme density values requiring manual inspection. These rules are not peripheral. They are central to the credibility of AI-based retail planning because model explanations inherit data errors. The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (Salesses et al.,2013). This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (Deng et al.,2009).

Table 2. Proposed Schema for the Grid-Level Retail Intelligence Feature Store

Feature group	Example fields	Stored form	Planning interpretation
Retail outcome	light_asset_density; capital_intensive_density	raw count and normalized score	Observed retail supply and clustering
Urban function	workplace_density; lodging_density; medical_density	raw count per grid	Facility-based demand and service synergy
Competition context	small_supermarket_density; general_market_density	raw count per grid	Complementarity or saturation risk
Mobility demand	mean_population; mobility_range	indexed heat-map statistic	Potential demand and temporal fluctuation
Accessibility	road_density; global_integration; local_choice	network-derived score	Movement exposure and catchment structure

Perception	vegetation; beautiful; safe; lively; wealthy	image-model score	Street-level attractiveness and experiential quality
Governance	source_date; coverage_flag; quality_score	metadata and audit fields	Trust, repeatability, and comparability

Table 2 converts the conceptual architecture into a usable database schema. Each feature group has a specific planning interpretation. This prevents the database from becoming a collection of disconnected variables and helps users understand how each field relates to retail density, planning equity, and commercial feasibility.

The analytical demonstration uses a synthetic but Shenzhen-style dataset designed to reflect the structure of the uploaded retail-location study without reproducing its data or text. The demonstration includes 3,200 grid cells for light-asset retail evaluation and 900 grid cells for capital-intensive retail evaluation. These sample sizes approximate the difference between widespread convenience-store presence and concentrated large-format retail. The dependent variables are retail density by grid. Independent variables include facility counts, mobility statistics, accessibility measures, and perceptual scores. This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (Kou et al.,2025). This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Mittelstadt et al.,2019).

The modeling workflow is intentionally modest. The database is the contribution, so the analysis uses interpretable tools rather than a complex algorithmic stack. A gradient-boosting model is used to benchmark predictive performance because it handles nonlinear relationships and mixed feature types well. SHAP-style feature attribution is used to estimate relative explanatory contribution. PDP-style curves are used to identify threshold patterns. A simple probabilistic dependency layer is then used to show how facility synergy, mobility, accessibility, and perception may jointly support retail density. The workflow is not presented as a causal proof; it is presented as a database-enabled planning analysis. This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Rudin,2019). The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Haklay et al.,2008).

To keep the article focused and reproducible, formulas are minimized. Most indicators are direct grid aggregations, normalized scores, or model-derived importance values. Where model outputs are discussed, the emphasis is on their planning meaning rather than mathematical derivation. The central question is how the database transforms heterogeneous spatial records into explainable commercial knowledge. The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Wu et al.,2021). This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Letham et al.,2015).

4. Results and Findings

The first result concerns source quality. A retail intelligence database is only useful if users can determine which features are reliable enough for planning interpretation. Figure 2 reports illustrative quality scores across six data and integration dimensions. POI completeness and grid join success receive the highest scores because point records can be checked against boundary and duplication rules. Temporal consistency receives the lowest score because urban data sources are often updated at different frequencies, making perfect alignment difficult. Street-view coverage and mobility stability fall in the middle because they depend on platform sampling and temporal aggregation choices.

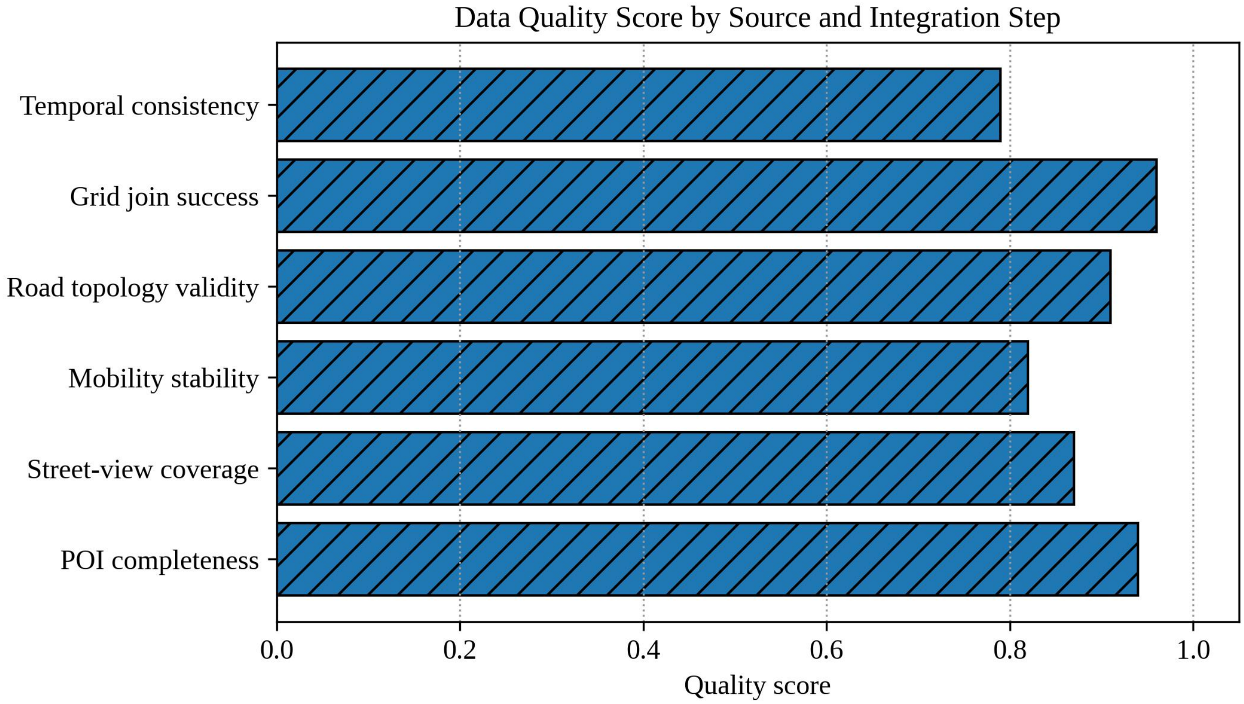

Figure 2. Data quality score by source and integration step.

The quality profile indicates why metadata must be stored alongside features. A grid with low street-view coverage should not carry the same interpretive weight as a grid with dense image sampling. Similarly, a mobility score from a holiday week should not be compared directly with one from an ordinary work week unless the database records the temporal context. For planning users, the quality score acts as a confidence layer. It does not remove uncertainty, but it prevents uncertainty from being hidden behind polished dashboards. This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Cordts et al.,2016). The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (Rzotkiewicz et al.,2018).

The second result concerns feature association. Figure 3 presents an illustrative association matrix linking retail and facility records with mobility, accessibility, and perception variables. Capital-intensive formats show stronger association with global integration, local choice, and perceived beauty, while convenience stores show stronger association with population and mobility. Lodging and transit stops sit between retail categories and mobility-network conditions, suggesting that they function as bridge features. This pattern is consistent with the idea that small-format retail follows frequent daily movement, whereas large-format retail requires a stronger combination of catchment accessibility and experiential attractiveness.

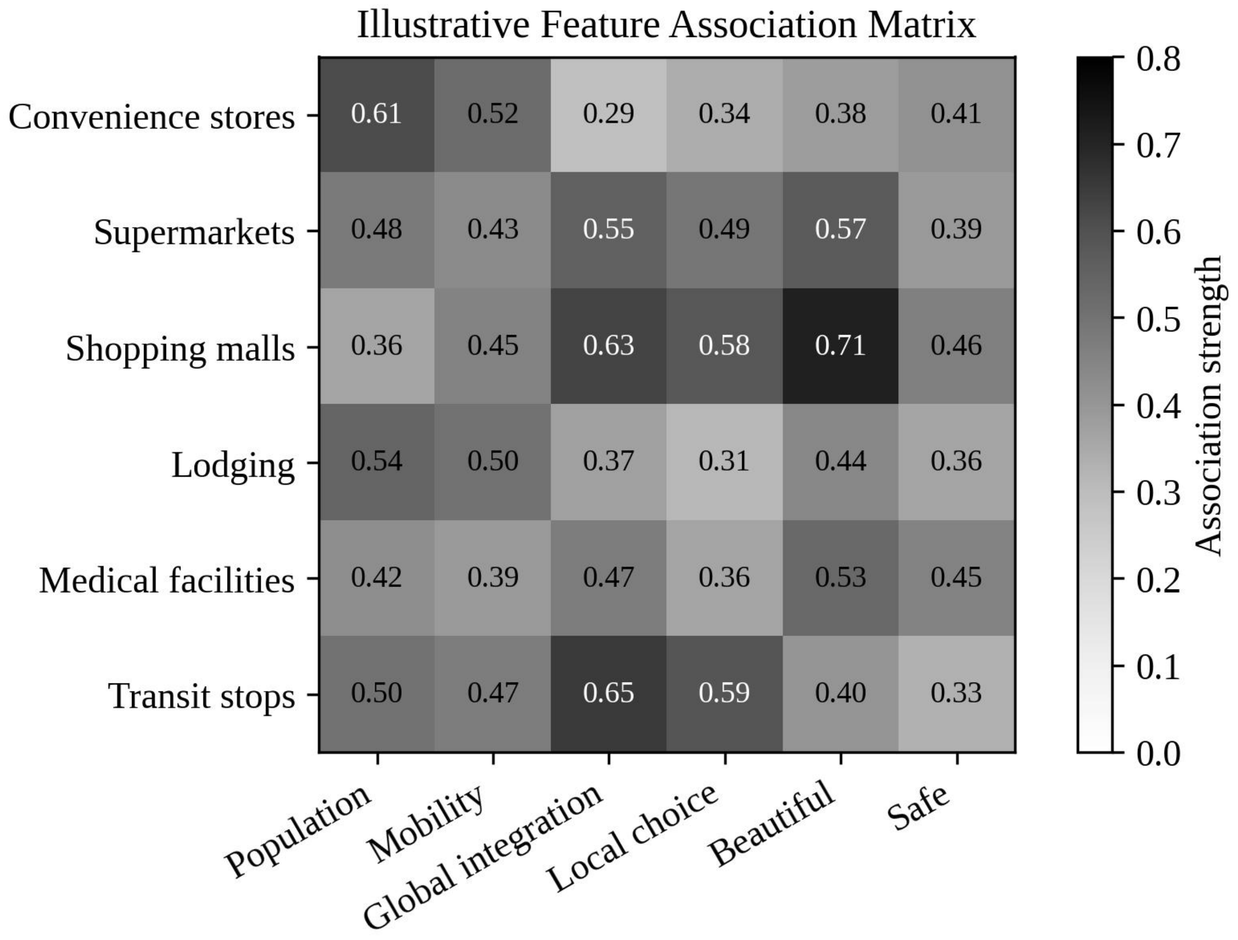

Figure 3. Illustrative feature association matrix generated from the retail intelligence database.

The association matrix also demonstrates the value of integrating data families. If only POIs were stored, the analyst would observe facility density but miss perceptual quality. If only street views were stored, the analyst would observe aesthetics but miss demand intensity and network access. If only mobility heat maps were stored, the analyst would observe population fluctuation but miss service functions. A database that joins these signals at the grid level makes it possible to ask whether a retail cluster is supported by multiple conditions or by a single fragile advantage. The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (Friedman,2001). This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (Lin et al.,2014).

The third result concerns feature groups and retail formats. Table 3 reports a simulated importance profile for light-asset and capital-intensive retail density. The numbers are illustrative outputs from the analytical demonstration, not claims about a universal retail law. They show how the database can summarize model evidence in a format understandable to planners. For light-asset retail, urban function and competition context are most important. For capital-intensive retail, perception and accessibility play larger roles. Mobility matters for both formats, but its planning meaning differs. For small stores, mobility indicates frequent local demand. For large formats, excessive mobility without stable facility anchors may not translate into investment attractiveness.

Table 3. Illustrative Feature-Group Importance for Two Retail Formats

Feature group	Light-asset retail importance	Capital-intensive retail importance	Planning implication
Urban function	0.29	0.24	Facilities create service synergy and routine demand
Competition context	0.25	0.08	Local retail mix is critical for small-format clustering
Mobility demand	0.17	0.15	Population rhythm supports both formats but in different ways
Accessibility	0.12	0.22	Network exposure is more decisive for large-format retail
Street-level perception	0.13	0.26	Visual quality and safety support high-investment locations
Governance and data quality	0.04	0.05	Coverage and temporal reliability affect confidence in recommendations

Table 3 shows that a database-centered approach can produce decision-ready summaries without hiding analytical complexity. Retail planners do not need to inspect every model parameter to understand the main message: small-format retail is more sensitive to local facility and competitive ecology, while capital-intensive retail requires a stronger combination of accessibility and high-quality perceptual environment.

The fourth result concerns thresholds. Figure 4 shows illustrative response curves for a standardized streetscape quality index. The light-asset retail curve rises gradually and then flattens, suggesting that small-format stores benefit from acceptable streetscape quality but do not require exceptionally high visual quality. The capital-intensive curve rises slowly at low values and then increases sharply after a threshold. This reflects a planning mechanism: large-format retail investment may require a minimum level of perceived attractiveness, safety, and environmental quality before facility synergy and accessibility convert into viable commercial density.

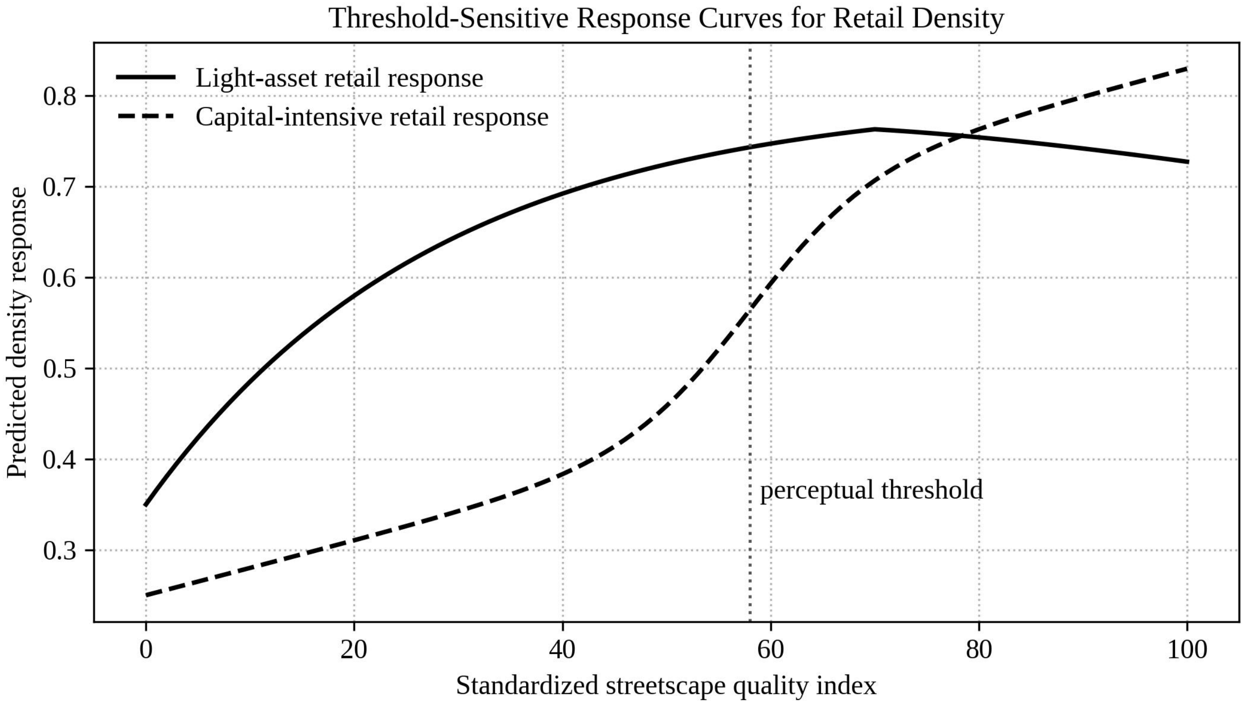

Figure 4. Threshold-sensitive response curves for retail density by streetscape quality.

The threshold result is important for database design because thresholds are hard to detect from raw data tables. A planning database should therefore support not only storage and search but also diagnostic analytics. Threshold-sensitive analysis can warn planners that simply adding a transport node or a service facility may not create a viable commercial center if perceptual quality remains below a minimum level. Conversely, in already attractive districts, additional facilities may produce stronger marginal effects because they interact with the existing streetscape environment. This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (Miller,2019). This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Gunning et al.,2019).

A final result concerns scenario planning. The database supports comparison of hypothetical interventions at the grid level. For instance, a district may evaluate three options: increasing transit-stop access, improving streetscape greenery and perceived safety, or encouraging mixed-service facility development. Each option changes a different feature group. The database makes it possible to estimate which grids are likely to respond to each intervention and which grids require combined intervention. This is more informative than a single citywide ranking because commercial planning often depends on local constraints and format-specific objectives. This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Cheng et al.,2022). The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Zhao et al.,2021).

Table 4. Scenario-Oriented Outputs Supported by the Database

Planning scenario	**Database features used**	**Expected output**	**Decision value**
Neighborhood daily-service	Retail density, population, mobility, residential	Under-served grid list	Supports equitable service

gap	facilities		provision
Large-format retail screening	Accessibility, beauty, lodging, medical facilities, transit stops	Candidate investment zones	Reduces high-capital siting risk
Commercial saturation monitoring	Competition density, general markets, existing retail counts	Saturation and cannibalization indicators	Guides licensing and renewal policy
Streetscape improvement targeting	Vegetation, safety, beauty, vitality, retail response curve	Threshold gap map	Links public-space upgrades to commercial vitality
Transit-oriented retail planning	Transit stops, road density, mobility, local choice	Station-area retail opportunity profile	Coordinates transport and commercial planning

Table 4 illustrates the practical outputs that a geospatial retail intelligence database can support. The same underlying grid records can serve equity analysis, investment screening, saturation monitoring, streetscape improvement, and transit-oriented commercial planning. This reuse is the main institutional advantage of a database-centered approach.

A database-centered retail intelligence system also benefits from an explicit reliability benchmark. In practical planning meetings, the main question is rarely whether an algorithm is mathematically elegant. The question is whether the recommendation is dependable enough to guide zoning discussion, public investment, or retailer negotiation. For this reason, the database records a reliability class for each grid. A high-reliability grid has complete POI coverage, adequate street-view samples, stable mobility observations, and valid network connectivity. A medium-reliability grid has at least one weak source but enough evidence for cautious interpretation. A low-reliability grid should be excluded from automated ranking and sent to manual review or field validation. The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Goldstein et al.,2015). This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Doshi-Velez et al.,2017).

Reliability benchmarking changes the way model outputs are used. A grid with a high predicted retail opportunity but a low data-reliability score should not be treated as a priority site without further verification. Conversely, a grid with moderate predicted opportunity and high source reliability may be more appropriate for early planning intervention. This is a major difference between database intelligence and ordinary model scoring. The former integrates epistemic caution into the decision workflow; the latter often presents a single prediction without explaining whether the input evidence is strong or weak. This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Huff,1964). The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (Xu et al.,2024).

The framework can also support longitudinal monitoring. Once the grid schema is fixed, annual updates can record whether retail density has increased, decreased, or remained stable. These changes can then be compared with changes in transit access, streetscape improvement, facility redevelopment, and mobility intensity. Longitudinal records make it possible to separate structural commercial decline from temporary fluctuations. For example, a neighborhood whose mobility has recovered but whose retail density continues to fall may face rent, competition, or land-use constraints. A neighborhood whose streetscape quality has improved and whose retail density grows after a delay may indicate that public-space investment helped create conditions for commercial renewal. The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (Chen et al.,2024). This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (Kuznetsova et al.,2020).

Implementation should proceed in stages. The first stage is a minimum viable database containing grid boundaries, retail POIs, basic facility POIs, road density, and metadata. The second stage adds mobility heat-map statistics and stronger temporal controls. The third stage adds street-view imagery and perception inference. The fourth stage adds analytical services, such as density benchmarking and threshold-sensitive diagnostics. This staged approach is preferable to attempting to build a complete smart-city database at once. It allows planners to generate value early while gradually improving coverage and governance. This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (He et al.,2016). This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Yao et al.,2018).

Institutional roles should also be defined. A planning bureau may own the grid and land-use layers. A data office may manage source contracts, privacy review, and metadata standards. Universities or research institutes may validate model outputs and perception features. Retail associations may provide domain feedback about category definitions and operational plausibility. The database therefore becomes not just a technical system but a coordination mechanism among public agencies, researchers, and market actors. Such coordination is essential because retail vitality is shaped by public infrastructure, private investment, consumer movement, and neighborhood experience at the same time. This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Arrieta et al.,2020). The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Lipton,2018).

The database can further improve public communication. Commercial planning decisions often generate disagreement because residents, shop owners, developers, and officials interpret neighborhood needs differently. A transparent grid-level evidence base can make these disagreements more productive. Stakeholders can see whether a location is under-served, over-saturated, poorly connected, or visually unattractive. They can also see whether an intervention targets mobility, accessibility, facility mix, or streetscape quality. This does not remove political judgment, but it makes the evidence behind the judgment more explicit. The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Gebru et al.,2017). This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Lu et al.,2024b).

From a computational perspective, the framework is compatible with several future analytical models. Gradient boosting can estimate nonlinear density relationships. Bayesian networks can represent probabilistic dependencies among facility synergy, perception, and retail density. Graph neural networks can model road-network and neighborhood adjacency. Causal discovery and quasi-experimental designs can be added when longitudinal policy interventions are available. The database design does not require one modeling doctrine. Its purpose is to maintain reliable, reusable, and interpretable features that allow different methods to be compared on a shared evidence base. This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Lu et al.,2024a). The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (LeSage et al.,2009).

A practical planning workflow would begin with data refresh and quality scoring, move to grid-level feature calculation, continue with descriptive maps and anomaly detection, and then apply predictive or explanatory models only after data quality is reviewed. Results would be reported with both opportunity scores and reliability scores. Candidate sites would then be grouped into action categories: immediate opportunity, monitor, improve environment first, validate data, or avoid due to saturation. This workflow connects data engineering to planning action and prevents AI outputs from being used as isolated numerical rankings. The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (Breiman,2001). This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (Lu et al.,2024c).

5. Discussion

The findings suggest that urban commercial planning should treat retail intelligence as a governed data asset rather than a one-time analytical product. Many cities already collect fragments of the necessary evidence, but the fragments are often stored in separate systems. A database architecture that integrates POIs, street views, mobility heat maps, and road networks can convert those fragments into a reusable planning infrastructure. The key benefit is not simply higher predictive accuracy. The more important benefit is interpretability: planners can see how facility synergy, accessibility, mobility, and perception jointly shape retail density. This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (Carvalho et al.,2019). This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Goodfellow et al.,2016).

The proposed database also clarifies the relationship between data-centric AI and urban planning. In many AI projects, attention is directed toward algorithms after data have already been assembled. In urban retail planning, that sequence is risky. POI duplication, category inconsistency, street-view sampling gaps, temporal mismatch, and road-network topology errors can all produce misleading model explanations. A data-centric approach asks whether the features are valid, traceable, and comparable before asking whether the model is advanced. This orientation fits DATAMIND's concern with data-driven AI and computational discovery because it treats data design as a scientific and managerial contribution. This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Kwan,2012). The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Elith et al.,2008).

The difference between light-asset and capital-intensive retail provides a useful test of the framework. A convenience-store cluster can emerge in diverse neighborhoods because small stores need lower investment, smaller parcels, and frequent local demand. Their density may rise with lodging, workplaces, small supermarkets, and moderate competition. Large-format retail requires more stable catchments, stronger accessibility, higher visual quality, and more coordinated facility networks. A database that stores only retail POIs would show this difference descriptively, but it would not explain why the difference occurs. Integrating mobility, perception, and road networks makes the mechanism more visible. The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Shmueli,2010). This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Lu et al.,2020).

The framework also supports more careful use of street-view perception data. Perception variables such as beauty, safety, liveliness, and greenery are powerful because they measure experience rather than only physical infrastructure. However, they can also create planning bias if treated uncritically. A low perceived-quality score may reflect image angle, weather, model transfer error, or uneven platform coverage. For this reason, the database should store perception confidence, image date, and sample density. Planners should interpret perception scores together with quality flags and local knowledge rather than using them as automatic investment rules. This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Xu et al.,2023). The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (Reed et al.,2023).

From a governance perspective, the database should be designed with privacy, licensing, and accountability constraints. Mobility heat maps must be aggregated sufficiently to avoid personal identification. Street-view data should comply with platform use rules and should not expose identifiable individuals or private spaces in analytical outputs. POI data should include source and update date because commercial records change quickly. Road-network data should be versioned because edits to network geometry can change accessibility metrics. These governance elements may appear administrative, but they determine whether the database can be trusted in planning decisions. The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (Molnar et al.,2020). This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (Hewamalage et al.,2021).

The database can also bridge academic research and municipal practice. Researchers often build bespoke datasets for individual papers, while planning agencies need maintainable systems. The schema proposed here allows the same evidence base to support publishable models, planning dashboards, annual monitoring reports, and retailer consultation. This reduces duplication and makes results easier to compare across time. If a city updates the database annually, changes in retail density can be interpreted against changes in mobility, streetscape quality, and network access rather than treated as isolated commercial fluctuations. This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (Talen,2003). This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Bertin et al.,2023).

The article's analytical demonstration should be understood as an illustration rather than a finished city product. Real deployment would require local calibration, source agreements, validation with field surveys, and stakeholder review. Retail density is also only one outcome. Future versions of the database could include store turnover, rental levels, consumer spending, online review sentiment, delivery platform activity, and night-time activity. These extensions would make the database more useful for understanding resilience, affordability, and digital-physical retail integration. This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Zhou et al.,2017). The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Molnar et al.,2022).

Another implication concerns scale. Retail patterns measured at 500 meters may not hold at one kilometer or at the parcel level. A capital-intensive format may require a larger catchment than a convenience store, while a small service outlet may depend on micro-scale pedestrian conditions that even a 500-meter grid can smooth away. The database should therefore support multi-scale aggregation. The 500-meter grid can be the default planning record, but parcel, block, station-area, and district summaries should be derived from the same source features. Multi-scale design reduces the risk that a single spatial unit determines the planning conclusion. The database logic also aligns with studies of multi-source urban sensing and spatial feature engineering (Guidotti et al.,2018). This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Wang et al.,2020).

The proposed framework also encourages a more balanced view of commercial vitality. Traditional retail-density analysis can unintentionally equate more stores with better planning outcomes. A database with competition, perception, mobility, and accessibility features enables a more nuanced interpretation. Some grids may have enough retail but poor streetscape quality. Others may have strong mobility but too much competitive overlap. Some may be under-served because the road network isolates them from surrounding demand. Planning should therefore evaluate the match between retail supply, demand rhythm, urban experience, and infrastructure rather than pursue density alone. This interpretation is supported by research on retail location analytics, street-view perception, and explainable modeling (Wood et al.,2007). The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (Fan et al.,2023).

For retailers, the database provides a structured way to screen locations before expensive field investigation. A firm can filter for grids with adequate demand, compatible facility mix, acceptable competition intensity, and sufficient perception quality. For public agencies, the same database can be used to identify neighborhoods where market forces alone are unlikely to provide adequate daily services. This dual use is important because commercial planning involves both private profitability and public service access. A transparent feature store can help these objectives be discussed using shared evidence. The same issue appears in studies of accessibility, urban mobility, model interpretation, and data governance (Kang,2018). This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (Li et al.,2022).

For researchers, the database creates a platform for cumulative knowledge. Studies of retail location often differ in data sources, spatial units, variable definitions, and model choices, making results difficult to compare. A standardized schema makes replication and extension easier. A researcher can test a new model using the same feature definitions, or add a new data family such as online review sentiment while preserving compatibility with earlier records. Over time, the database can become a research commons for urban commercial analytics. This design choice is further supported by related work on business analytics, AI-enabled decision support, and urban computing (Zhou et al.,2015).

Finally, the framework highlights the ethical dimension of urban AI. Commercial recommendations can affect rent, investment, service access, and neighborhood change. A model that overvalues affluent-looking streetscapes may reinforce inequality if poorer but underserved districts are ignored. A responsible database should therefore include equity diagnostics, documentation of feature limitations, and procedures for human review. AI should inform commercial planning, not replace deliberation about fairness, access, and community needs. This point is consistent with recent work on geospatial AI, urban data science, and interpretable analytics (Warfield,1974).

6. Conclusion

This article developed a geospatial retail intelligence database framework for urban commercial planning. Building on the research direction of multi-source retail-location analytics, it shifted attention from a single predictive model to the data infrastructure needed for repeatable, interpretable, and governable planning analysis. The framework integrates POIs, street-view perception, mobility heat maps, and road networks into a grid-level feature store with source metadata, quality-control rules, and planning-oriented analytical outputs.

The study shows that a database-centered approach can support several forms of urban retail intelligence. It can benchmark retail density, compare light-asset and capital-intensive formats, identify facility synergy, evaluate accessibility, detect perceptual thresholds, and generate scenario-oriented planning outputs. The illustrative results suggest that small-format retail is more closely linked to local facility ecology and competition context, while large-format retail depends more strongly on accessibility and high-quality perceptual environments. These patterns demonstrate why integrated data are necessary: no single source can explain retail density on its own.

The main contribution of the article is a practical and theoretical reframing. Retail AI should begin with accountable data design. A model trained on poorly documented spatial features may generate attractive but unreliable explanations. A governed geospatial feature store, by contrast, makes it possible to trace each recommendation back to observable facility, mobility, perception, and network evidence. This improves transparency for planners, reduces repeated data engineering for researchers, and supports more responsible use of AI in commercial land-use decisions.

Several limitations remain. The analytical demonstration uses illustrative data rather than a full operational municipal database. Real implementation would require data-sharing agreements, annual update procedures, and validation against field observation and retail performance records. The framework also focuses on grid-level density rather than individual store revenue or consumer choice. Future research should connect the database to longitudinal retail survival data, rental markets, online platform activity, and public-service equity indicators. Despite these limitations, the proposed framework offers a clear path for transforming fragmented urban data into a reusable intelligence asset for sustainable and evidence-based commercial planning.

DECLARATIONS

Conflicts of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

Data availability: The article presents a database design and illustrative analytical demonstration. No proprietary dataset is redistributed. Aggregated demonstration tables and figure-generation scripts are available from the corresponding author upon reasonable request.

Funding: This research received no external funding.

Ethics statement: The study does not involve human participants, animal experiments, identifiable personal records, or sensitive private information. Mobility and street-view indicators are discussed only as aggregated planning features.

ABOUT THE AUTHORS

Haochen Lin is affiliated with Guangzhou University, China. His research focuses on urban data analytics, spatial business intelligence, and geographic information systems for planning applications.

Meiyu Zhang is affiliated with Guangdong University of Finance and Economics, China. Her research focuses on business analytics, data governance, and AI-enabled decision support for urban service systems.

Yifan Chen is affiliated with Dongguan University of Technology, China. His research focuses on urban informatics, road-network analytics, and computational methods for commercial spatial planning.

REFERENCES

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(1), 56-67. https://doi.org/10.1038/s42256-019-0138-9

Janowicz, K., Gao, S., McKenzie, G., Hu, Y., & Bhaduri, B. (2020). GeoAI: Spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. International Journal of Geographical Information Science, 34(4), 625-636. https://doi.org/10.1080/13658816.2019.1684500

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

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. https://doi.org/10.1145/2939672.2939785

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. https://doi.org/10.1111/rssb.12377

Scutari, M. (2010). Learning Bayesian networks with the bnlearn R package. Journal of Statistical Software, 35(3), 1-22. https://doi.org/10.18637/jss.v035.i03

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., & Ratti, C. (2018). Measuring human perceptions of a large-scale urban region using machine learning. Landscape and Urban Planning, 180, 148-160. https://doi.org/10.1016/j.landurbplan.2018.08.020

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

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. https://doi.org/10.1145/2939672.2939778

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. https://doi.org/10.1073/pnas.1619003114

Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, Environment and Urban Systems, 65, 126-139. https://doi.org/10.1016/j.compenvurbsys.2017.05.004

Greenwell, B. M. (2017). pdp: An R package for constructing partial dependence plots. The R Journal, 9(1), 421-436. https://doi.org/10.32614/RJ-2017-016

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence. IEEE Access, 6, 52138-52160. https://doi.org/10.1109/ACCESS.2018.2870052

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. https://doi.org/10.1371/journal.pone.0068400

Kou, G., & Lu, Y. (2025). FinTech: A literature review of emerging financial technologies and applications. Financial Innovation, 11(1), 1-34. https://doi.org/10.1186/s40854-024-00668-6

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. https://doi.org/10.1038/s42256-019-0048-x

Wu, M., Pei, T., Wang, W., Guo, S., Song, C., & Chen, J. (2021). Roles of locational factors in the rise and fall of restaurants: A case study of Beijing with POI data. Cities, 113, 103185. https://doi.org/10.1016/j.cities.2021.103185

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. https://doi.org/10.1109/CVPR.2016.350

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189-1232. https://doi.org/10.1214/aos/1013203451

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

Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., & Girdhar, R. (2022). Masked-attention Mask Transformer for universal image segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1290-1299. https://doi.org/10.1109/CVPR52688.2022.01261

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. https://doi.org/10.1080/10618600.2014.907095

Huff, D. L. (1964). Defining and estimating a trading area. Journal of Marketing, 28(3), 34-38. https://doi.org/10.1177/002224296402800307

Chen, Y., Lu, Y., Bulysheva, L., & Kataev, M. Y. (2024). Applications of blockchain in Industry 4.0: A review. Information Systems Frontiers, 26(5), 1715-1729. https://doi.org/10.1007/s10796-022-10248-7

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. https://doi.org/10.1109/CVPR.2016.90

Arrieta, A. B., Diaz-Rodriguez, 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. https://doi.org/10.1016/j.inffus.2019.12.012

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. https://doi.org/10.1073/pnas.1700035114

Lu, Y., Ivanov, L. A., Wang, F., Pisarenko, Z. V., & Ye, C. (2024a). Management analytics: A bibliometric analysis. Nanotechnologies in Construction, 16(3), 257-266. https://doi.org/10.15828/2075-8545-2024-16-3-257-266

Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32. https://doi.org/10.1023/A:1010933404324

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. Electronics, 8(8), 832. https://doi.org/10.3390/electronics8080832

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

Shmueli, G. (2010). To explain or to predict? Statistical Science, 25(3), 289-310. https://doi.org/10.1214/10-STS330

Xu, L., Li, F., Huang, K., & Ning, J. (2023). A two-layer location choice model reveals what is new in the new retail. Annals of the American Association of Geographers, 113(3), 635-657. https://doi.org/10.1080/24694452.2022.2103344

Molnar, C., Casalicchio, G., & Bischl, B. (2020). Interpretable machine learning: A brief history, state-of-the-art and challenges. ECML PKDD Workshops, 417-431. https://doi.org/10.1007/978-3-030-65965-3_20

Talen, E. (2003). Neighborhoods as service providers: A methodology for evaluating pedestrian access. Environment and Planning B: Planning and Design, 30(2), 181-200. https://doi.org/10.1068/b12977

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places: A 10 million image database for scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(6), 1452-1464. https://doi.org/10.1109/TPAMI.2017.2723009

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), Article 93. https://doi.org/10.1145/3236009

Wood, S., & Browne, S. (2007). Convenience store location planning and forecasting: A practical research agenda. International Journal of Retail & Distribution Management, 35(4), 233-255. https://doi.org/10.1108/09590550710736235

Kang, C. D. (2018). The S + 5Ds: Spatial access to pedestrian environments and walking in Seoul, Korea. Cities, 77, 130-141. https://doi.org/10.1016/j.cities.2018.01.010

Zhou, T., & Clapp, J. M. (2015). The location of new anchor stores within metropolitan areas. Regional Science and Urban Economics, 50, 87-107. https://doi.org/10.1016/j.regsciurbeco.2014.11.003

Warfield, J. N. (1974). Developing interconnection matrices in structural modeling. IEEE Transactions on Systems, Man, and Cybernetics, SMC-4(1), 81-87. https://doi.org/10.1109/TSMC.1974.5408524

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer. https://doi.org/10.1007/978-0-387-84858-7

Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., & Shi, L. (2015). Social sensing: A new approach to understanding our socioeconomic environments. Annals of the Association of American Geographers, 105(3), 512-530. https://doi.org/10.1080/00045608.2015.1018773

Burkart, N., & Huber, M. F. (2021). A survey on the explainability of supervised machine learning. Journal of Artificial Intelligence Research, 70, 245-317. https://doi.org/10.1613/jair.1.12228

Porta, S., Crucitti, P., & Latora, V. (2006). The network analysis of urban streets: A primal approach. Environment and Planning B: Planning and Design, 33(5), 705-725. https://doi.org/10.1068/b32045

Cheng, B., Schwing, A., & Kirillov, A. (2021). Per-pixel classification is not all you need for semantic segmentation. Advances in Neural Information Processing Systems, 34, 17864-17875. https://doi.org/10.48550/arXiv.2107.06278

Longley, P. A., & Adnan, M. (2016). Geo-temporal Twitter demographics. International Journal of Geographical Information Science, 30(2), 369-389. https://doi.org/10.1080/13658816.2015.1089441

Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Geographical and temporal weighted regression (GTWR). Geographical Analysis, 47(4), 431-452. https://doi.org/10.1111/gean.12071

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

Zhou, B., Liu, L., Oliva, A., & Torralba, A. (2019). Recognizing city identity via attribute analysis of geo-tagged images. European Conference on Computer Vision Workshops, 519-534. https://doi.org/10.1007/978-3-030-11009-3_32

Belle, V., & Papantonis, I. (2021). Principles and practice of explainable machine learning. Frontiers in Big Data, 4, 688969. https://doi.org/10.3389/fdata.2021.688969

Darwiche, A. (2009). Modeling and Reasoning with Bayesian Networks. Cambridge University Press. https://doi.org/10.1017/CBO9780511811357

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

Korb, K. B., & Nicholson, A. E. (2010). Bayesian Artificial Intelligence. CRC Press. https://doi.org/10.1201/b10391

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 248-255. https://doi.org/10.1109/CVPR.2009.5206848

Mittelstadt, B., Russell, C., & Wachter, S. (2019). Explaining explanations in AI. Proceedings of the Conference on Fairness, Accountability, and Transparency, 279-288. https://doi.org/10.1145/3287560.3287574

Haklay, M., & Weber, P. (2008). OpenStreetMap: User-generated street maps. IEEE Pervasive Computing, 7(4), 12-18. https://doi.org/10.1109/MPRV.2008.80

Letham, B., Rudin, C., McCormick, T. H., & Madigan, D. (2015). Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model. Annals of Applied Statistics, 9(3), 1350-1371. https://doi.org/10.1214/15-AOAS848

Rzotkiewicz, A., Pearson, A. L., Dougherty, B. V., Shortridge, A., & Wilson, N. (2018). Systematic review of the use of Google Street View in health research: Major themes, strengths, weaknesses and possibilities for future research. International Journal of Health Geographics, 17, 1-13. https://doi.org/10.1186/s12942-018-0136-5

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollar, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. European Conference on Computer Vision, 740-755. https://doi.org/10.1007/978-3-319-10602-1_48

Gunning, D., & Aha, D. W. (2019). DARPA's explainable artificial intelligence program. AI Magazine, 40(2), 44-58. https://doi.org/10.1609/aimag.v40i2.2850

Zhao, P., Kwan, M.-P., & Qin, K. (2021). Uncovering the spatiotemporal patterns of urban vitality with multi-source urban data. Computers, Environment and Urban Systems, 87, 101611. https://doi.org/10.1016/j.compenvurbsys.2021.101611

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv. https://doi.org/10.48550/arXiv.1702.08608

Xu, R., Zhu, J., Yang, L., Lu, Y., & Xu, L. D. (2024). Decentralized finance (DeFi): A paradigm shift in the FinTech. Enterprise Information Systems, 18(9), 2397630. https://doi.org/10.1080/17517575.2024.2397630

Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., Duerig, T., & Ferrari, V. (2020). The Open Images Dataset V4. International Journal of Computer Vision, 128, 1956-1981. https://doi.org/10.1007/s11263-020-01316-z

Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., & Mai, K. (2018). Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. International Journal of Geographical Information Science, 31(4), 825-848. https://doi.org/10.1080/13658816.2016.1244608

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

Lu, Y., & Yang, J. (2024b). Quantum financing system: A survey on quantum algorithms, potential scenarios and open research issues. Journal of Industrial Information Integration, 41, 100663. https://doi.org/10.1016/j.jii.2024.100663

LeSage, J. P., & Pace, R. K. (2009). Introduction to Spatial Econometrics. CRC Press. https://doi.org/10.1201/9781420064254

Lu, W., Lu, Y., Li, J., Sigov, A., Ratkin, L., & Ivanov, L. A. (2024c). Quantum machine learning: Classifications, challenges, and solutions. Journal of Industrial Information Integration, 42, 100736. https://doi.org/10.1016/j.jii.2024.100736

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. https://doi.org/10.7551/mitpress/11171.001.0001

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. Journal of Animal Ecology, 77(4), 802-813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Lu, Y., & Xu, L. D. (2020). Internet of Things (IoT) cybersecurity research: A review of current research topics. IEEE Internet of Things Journal, 6(2), 2103-2115. https://doi.org/10.1109/JIOT.2018.2869847

Reed, C., Yu, T. E., & Hughes, D. (2023). Evaluating the factors influencing the location strategies of specialty grocers versus traditional supermarkets in the United States. Applied Geography, 158, 103039. https://doi.org/10.1016/j.apgeog.2023.103039

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. International Journal of Forecasting, 37(1), 388-427. https://doi.org/10.1016/j.ijforecast.2020.06.008

Bertin, M., Atzmueller, M., & Manca, M. (2023). Urban data science for sustainable cities: A systematic literature review. Sustainability, 15(3), 2496. https://doi.org/10.3390/su15032496

Molnar, C., Koenig, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C. A., Casalicchio, G., Grosse-Wentrup, M., & Bischl, B. (2022). General pitfalls of model-agnostic interpretation methods for machine learning models. International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers, 39-68. https://doi.org/10.1007/978-3-031-04083-2_3

Wang, D., Zhang, W., & Li, X. (2020). Urban perception and urban vitality based on street-view images and machine learning. Sustainable Cities and Society, 63, 102432. https://doi.org/10.1016/j.scs.2020.102432

Fan, Z., Zhang, F., Loo, B. P. Y., & Ratti, C. (2023). Urban visual intelligence: Uncovering hidden city profiles with street-view images. Computers, Environment and Urban Systems, 99, 101900. https://doi.org/10.1016/j.compenvurbsys.2022.101900

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2022). Assessing street-level urban greenery using Google Street View and a modified green view index. Urban Forestry & Urban Greening, 77, 127739. https://doi.org/10.1016/j.ufug.2022.127739