UNEC Tələbə Tədqiqatları Jurnalı, 3(1), 2026

DOİ: https://doi.org/10.30546/301678.01.010.2026.573

# AI-ASSISTED GEOGRAPHIC ASSESMENT OF BUSINESS LOCATIONS

**Baghirova Fidan Khayyam**

*Azerbaijan State University of Economics (UNEC)*

*Faculty of Digital Economy, class 574-i*

Email: bagirova.fidan.xayyam.2024@unec.edu.az

**Scientific Supervisor: Roza Misirli**

*Azerbaijan State University of Economics (UNEC), Baku, Azerbaijan*

Email: roza-misirli@unec.edu.az

## ABSTRACT

This article presents an artificial intelligence-assisted geographic decision-support prototype for evaluating potential business locations in Azerbaijan. The study combines a local GeoJSON dataset of 11,807 objects and 69 object types, district-level demographic indicators for Baku, and a Python-based spatial scoring model implemented in a Streamlit application. A large language model is used only to transform a natural-language business idea into structured intent, including target business categories and supporting facility types. The final location recommendation is not generated by the language model, it is produced through deterministic calculations based on Haversine distance, competitor density, supporting-object availability, local demand proxy and demographic advantage. The prototype groups nearby coordinate into grid-based candidate areas, ranks them by a transparent multi-factor score, and presents the results with map visualization and explanatory evidence. The study demonstrates that semantic AI can strengthen business-location analysis when it is combined with verifiable geographic data and explainable mathematical logic.

**Keywords:** *artificial intelligence, business location, geographic information systems, Haversine distance, GeoJSON, location scoring, decision support*

**JEL classification codes:** ***C63, C88, L81, R12, R32***

## INTRODUCTION

Location choice is one of the most consequential strategic decisions for urban businesses. In retail, food service, pharmacies, personal services and other consumer-facing activities, the chosen site affects customer accessibility, competitive pressure, operating cost, visibility and long-term revenue potential. In practice, small and medium-sized entrepreneurs often rely on intuition, rent level, street familiarity or informal observation when selecting a location. These factors are useful, but they do not provide a systematic way to compare alternatives across distance, nearby demand generators, competitor density and district-level socio-economic conditions.

Geographic information science provides the conceptual and technical foundation for transforming such decisions into measurable spatial problems. Goodchild (1992: 31) argues that geographic information science should be understood not merely as map production, but as a scientific field concerned with the representation, analysis and use of spatial information. In a related tradition, location science examines how facilities should be placed in relation to demand, distance and service coverage. Church (2002: 541) emphasizes that geographic information systems make location problems more operational because they connect spatial data, analytical models and decision support. The present article develops this logic for a practical business-location prototype. The research problem is the following: can a natural-language business idea be connected with real geographic objects, distance calculations, competitor density, supporting facilities and demographic indicators in order to produce explainable location recommendations? The proposed answer is a hybrid system in which artificial intelligence is used for semantic interpretation, while the actual spatial recommendation is generated by deterministic calculations over real coordinates.

The contribution of the study is therefore not the use of a language model as an independent decision maker. On the contrary, the system limits the role of language models. It extracts the likely business category and supports object types from user input, but it does not invent coordinates, assign arbitrary scores or select locations by textual reasoning. The ranking is calculated in a separate geographic logic module. This design is important because business-location recommendations must be interpretable, reproducible and open to verification.

**Literature Review**

Business-location analysis has long been related to trade-area theory and spatial consumer behavior. Huff (1964: 34) showed that trade areas cannot be understood only by administrative boundaries or simple distance. Instead, consumer attraction depends on the relationship between facility attractiveness, distance and the spatial distribution of alternatives. Although the present prototype does not implement the full Huff model, the same conceptual principle is used: a business site is evaluated not as an isolated point, but as part of a surrounding urban environment.

GIS-based location analysis extends this tradition by turning spatial relationships into computable procedures. Church (2002:542) explains that GIS and location science are complementary because GIS stores and visualizes spatial data while location science supplies models for facility placement, service accessibility and spatial optimization. This combination is directly reflected in the prototype. The application visualizes object density on a map, but the recommendation mechanism depends on calculated relationships between candidate sites and nearby facilities.

The data layer of the project is based on GeoJSON, a widely used format for encoding geographic features. RFC 7946 defines GeoJSON as a geospatial data interchange format based on JSON objects, where geographic coordinates and feature properties are stored together (Butler et al. 2016). This format is suitable for the present study because each object can be processed both as a map point and as an analytical record containing city, district, street, name and type information.

Open and user-generated mapping data have also expanded the possibilities of local spatial analysis. Haklay and Weber (2008: 12) discuss OpenStreetMap as an example of participatory geographic data production and show how collaborative mapping can support urban and spatial applications. The dataset used in this project is stored locally rather than queried live from OpenStreetMap, but the analytical idea is similar: structured point-of-interest data can be converted into evidence for practical urban decisions (Sadigov, 2025).

Distance measurement is a core technical element in location assessment. When locations are represented by latitude and longitude, Euclidean distance is not appropriate unless the area is very small and projected coordinates are used. For geographic coordinates, the Haversine formula provides a practical way to estimate great-circle distance between two points on the Earth. Sinnott (1984: 159) describes the formula as numerically stable and useful for calculating angular distance. In this prototype, Haversine distance is used to identify competitors and supporting facilities within a selected radius.

The artificial intelligence component is grounded in the ability of modern language models to interpret natural-language instructions. The Transformer architecture introduced by Vaswani et al. (2017: 5998) made attention mechanisms central to modern language processing. Later, Brown et al. (2020: 1877) demonstrated that large language models can perform a broad range of language tasks from limited examples. In the present study, this capability is used narrowly and cautiously: the model converts an informal business idea into a structured JSON object that can be processed by the deterministic spatial module.

Finally, the implementation depends on the scientific Python ecosystem. Pandas supports tabular data cleaning, filtering and grouping, which are necessary for preparing the GeoJSON-derived records and demographic table (McKinney 2010: 56). NumPy supports efficient array operations and vectorized mathematical calculation, including the distance computations used by the geographic module (Harris et al. 2020: 357). These tools make it possible to evaluate thousands of candidate locations without relying on manual inspection.

**Data and Technological Structure**

The prototype consists of four main components: a Streamlit user interface, a GeoJSON data-processing layer, a large-language-model interpretation layer and a deterministic geographic scoring module. The main application file is app.py. It loads datasets, creates filters for city, district and object type, visualizes points on a map and provides an input field for a natural-language business idea. The supporting functions are in core_utils.py, llm_agent.py and geo_logic.py.

The primary spatial dataset is dataset/azrbaycanda_obyektlr_v_mkanlar.geojson. Each feature includes coordinate geometry and attribute fields such as CITY, STATE, STRT_NAME, STRT_ADDR, NAME and TYPE. After parsing the file, the application obtains 11,807 valid point objects. These objects cover three main urban areas in the dataset, namely Baku, Absheron and Sumgayit, and include 69 object types. The most frequent categories include grocery stores, restaurants and cafes, personal services, schools, banks, hospitals, pharmacies and retail facilities.

The second dataset is dataset/baku_demographics.csv. It contains district-level indicators for Baku, including population density, number of employees and average monthly salary. These variables are used as socio-economic proxies. Population density approximates potential local customer volume; employment indicates daily economic activity and commuting-related demand; average salary approximates purchasing power. Since the demographic data are available at district level, they are used as an additional adjustment rather than as a micro-location measure.

The visualization layer is implemented with PyDeck inside Streamlit. The application uses a Google Maps tile layer as the base map, a heatmap layer to show spatial density and a scatterplot layer to show individual objects. This gives the user two complementary views: a general picture of object concentration and precise point-level information for specific facilities.

The language-model layer is implemented in llm_agent.py through an API request to a chat-completion endpoint. The model is prompted to return a JSON object with three main elements: the target business type, alternative target types and supporting facility types. For example, if the user writes that they want to open an affordable restaurant near students and transport stops, the model should identify restaurants as the target category and schools, universities, bus stops or similar categories as supporting facilities. The prompt explicitly instructs the model not to generate coordinates. The scoring layer is implemented in geo_logic.py. This module receives the resolved target and supporting object types, the selected search radius and the required number of top results. It then evaluates candidate coordinate groups, counts nearby competitors and supporting facilities, calculates a demand proxy and adds a demographic adjustment where district data are available.

**Methodology**

The methodology has three stages: semantic intent extraction, spatial candidate construction and multi-factor scoring. In the first stage, the user submits a business idea in natural language. The language model parses the text and returns a structured representation. This stage solves a practical usability problem. Entrepreneurs normally do not think in dataset category names such as

RESTAURANT_CAFES, PHARMACY or GROCERY_STORE. They describe a business goal in ordinary language. The model bridges this gap by mapping the user's wording to likely business and support categories.

In the second stage, the extracted categories are matched to real object types in the dataset. The application normalizes labels by converting them to uppercase, replacing non-alphanumeric characters and comparing the normalized form with known dataset types. This reduces the risk that a useful category will be missed because of small differences in wording.

In the third stage, the system constructs candidate locations. Instead of evaluating every object as a separate site, the module rounds coordinates to three decimal places:

$$lat\_grid = round(latitude, 3)$$
$$lon\_grid = round(longitude, 3)$$

This creates compact spatial groups of nearby points. In Azerbaijan's latitude range, three decimal places correspond roughly to about 100-110 meters in latitude, with longitude distance varying by latitude. This grid-based approach is appropriate for a prototype because a new business is not usually placed exactly on an existing object's coordinate. The system therefore evaluates small urban areas rather than individual existing facilities.

For each candidate grid, the module identifies surrounding objects within the user-selected radius. To improve efficiency, it first creates a bounding box around the candidate:

$$lat\_delta = radius\_km / 110.574$$
$$lon\_delta = radius\_km / (111.320 * cos(latitude))$$

This preliminary filter selects only points that could plausibly fall within the radius. The module then applies the Haversine formula for precise distance calculation:

$$a = sin\^{}2(Delta\ phi / 2) + cos(phi\_1) * cos(phi\_2) * sin\^{}2(Delta\ lambda / 2)$$
$$c = 2 * atan2(sqrt(a), sqrt(1 - a))$$
$$distance = R * c$$

Where R = 6371 km is the approximate radius of the Earth. A point is treated as part of the candidate's surrounding environment when its calculated distance is less than or equal to the selected radius.

Competitor density is calculated as the number of target-type objects within the radius:

$$competitors = count(target\text{-}type\ objects\ within\ radius)$$

A higher competitor count reduces the score. However, low competition alone is not treated as sufficient evidence of a good location. A location with no competitors may also have no customer-generating facilities. For this reason, the model separately calculates support strength:

$$key\_total = count(all\ supporting\ objects\ within\ radius)$$
$$key\_type\_coverage = count(supporting\ object\ types\ present\ within\ radius)$$
$$support\_density = key\_total / object\_count\_in\_grid$$

The model also includes a demand proxy based on the number of objects in the candidate grid:

$$demand\_proxy = \log(1 + object\_count\_in\_grid)$$

The logarithmic form prevents very dense grids from dominating the score excessively. It represents general urban activity rather than direct customer demand.

The demographic component is calculated from normalized district-level indicators:

$$demo\_score = (Population\_Density / max\_density) \times 1.5 + (Average\_Salary / max\_salary) \times 2.0 + (Employees / max\_employees) \times 0.5$$

Average salary receives the highest weight because it approximates purchasing power. Population density receives the second-highest weight because it reflects the potential local customer base. Employment receives smaller weight because it reflects daytime activity but may not always translate into consumer spending for every business type. The final location score is calculated as follows:

$$score = 4.0 \times support\_density + 2.5 \times key\_type\_coverage + 0.5 \times demand\_proxy - 1.75 \times competitors + demo\_boost$$

The formula combines five factors: density of supporting facilities, diversity of supporting facility types, general local activity, competitor pressure and demographic advantage. The score is not intended to be a universal measure of profitability. It is a decision-support ranking that helps compare candidate areas under transparent assumptions.

After ranking, the prototype applies to a simple diversity mechanism. It avoids returning top results that are all located on the same street or within 0.75 kilometers of one another when alternatives are available. This improves the usefulness of the output because users receive geographically distinct options rather than repeated variations of the same micro-area.

**Results and Interpretation**

The implemented prototype produces two types of output: an interactive spatial dashboard and an analytical recommendation report. The dashboard allows the user to filter the dataset by city, district and object type. It displays the total number of dataset records, the number of filtered objects and the number of available object categories. The map visualization provides a practical overview of spatial concentration. Dense clusters can be identified through the heatmap, while individual points remain available through the scatterplot layer and tooltip information.

The recommendation workflow begins when the user enters a business idea and selects a search radius and the number of top suggestions. The language model returns target and support categories. The deterministic logic module then evaluates all candidate grids in the selected scope. For each top location, the system reports city, district, street name, coordinates, Google Maps link, competitor count, total number of supporting objects, support-type coverage, demand proxy, score and nearby examples of supporting facilities (Amanova et al.,2026). This structure is important because it makes the recommendation explainable. The user does not only receive a statement that one location is "good". The system shows why the location was ranked highly. For instance, a candidate area may receive a strong score because it has several schools and transport-related objects nearby, a moderate level of existing competitors and favorable demographic conditions. Another candidate may score lower because, although competition is weak, the surrounding support objects are also limited. The results also demonstrate the advantage of separating semantic interpretation from spatial computation. The language model can understand user intent, but the candidate coordinates and scores are determined only after the real dataset is processed. This reduces the risk of hallucinated locations and creates an audit trail: every recommendation can be traced back to object counts, distances and demographic variables (Mammadov et al., 2025; Mustafayev et al, 2025).

From a practical business perspective, the prototype supports early-stage location screening. A small entrepreneur can compare alternative districts and streets before investing time in field visits, rent negotiations or market surveys. The tool does not replace professional feasibility analysis, but it can reduce the search space and make the first stage of decision-making more evidence based.

**Scientific Novelty and Practical Significance**

The scientific novelty of the study lies in the controlled integration of natural-language AI with deterministic GIS-based scoring. Many AI applications treat the model's generated response as the final output. In this prototype, the language model is deliberately restricted to semantic interpretation and explanation. The location score is produced by transparent mathematical operations over structured data. This architecture improves reliability because the most consequential decision, the ranking of locations, is not left to generative text. The practical significance is also clear. Small and medium-sized businesses often do not have access to advanced market-intelligence platforms. A lightweight Streamlit application using local datasets can provide an accessible alternative for preliminary site assessment. The same approach can be adapted for restaurants, cafes, pharmacies, grocery stores, service businesses and other urban activities. Municipal institutions and urban analysts could also use a similar method to examine service coverage, facility concentration and spatial inequality in access to important services.

**Limitations**

The findings depend on the completeness and freshness of the dataset. If the GeoJSON file contains outdated, missing or incorrectly classified objects, the recommendations may be affected. The model also assumes that existing point-of-interest density is a useful proxy for local activity, but this proxy does not directly measure pedestrian traffic, vehicle traffic, rental price, visibility, parking availability or legal zoning. The demographic data are available only at district level. This creates a limitation for micro-location analysis because two streets in the same district may differ significantly in income profile, footfall, accessibility and commercial value. In addition, the scoring weights are expert-defined rather than estimated from historical business performance data. Future research could improve the model by calibrating weights against real business outcomes such as revenue, closure rates or customer volume (Sadigov et al., 2025; Sultan et al., 2025). Another limitation is that the prototype evaluates circular catchment areas based on geographic distance. In real urban conditions, travel time, street networks, public transport routes and physical barriers may be more important than straight-line distance. Adding network-based travel-time analysis would make the model more realistic.

## RESULTS

This article presented an AI-assisted geographic prototype for evaluating business locations in Azerbaijan. The system processes a GeoJSON dataset of 11,807 objects, integrates Baku district-level demographic indicators, interprets natural-language business ideas through a large language model and ranks candidate areas through deterministic spatial scoring. The model combines support density, support-type diversity, demand proxy, competitor pressure and demographic advantage into an explainable location score.

The study shows that large language models can be useful in business-location analysis when their role is carefully limited. They can translate human business ideas into structured analytical categories, while the final recommendation remains grounded in real coordinates and verifiable calculations. This separation between semantic AI and mathematical GIS logic makes the system more transparent, more reproducible and more suitable for decision support.

Future development should extend the model with rental prices, pedestrian and vehicle traffic, public transport accessibility, road-network travel times, zoning restrictions and historical business-performance data. These additions would allow the prototype to move from preliminary location screening toward stronger prediction and optimization.

## REFERENCES


1. Amanova SHS və b. (2026). "Vegetation development and productivity in agrolandscapes using satellite imagery and ground surveys". *SABRAO Journal of Breeding and Genetics*, Vol. 58, No. 2, pp. 762-770. http://doi.org/10.54910/sabrao2026.58.2.26.
2. Brown Thomas B. və b. (2020). "Language Models are Few-Shot Learners". *Advances in Neural Information Processing Systems*, Vol. 33, pp. 1877-1901. https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html
3. Butler Howard (2016). *The GeoJSON Format*. RFC 7946. Internet Engineering Task Force. https://doi.org/10.17487/RFC7946. https://www.rfc-editor.org/rfc/rfc7946
4. Church Richard L. (2002). "Geographical Information Systems and Location Science". *Computers & Operations Research*, Vol. 29, No. 6, pp. 541-562. https://doi.org/10.1016/S0305-0548(99)00104-5.
5. Goodchild Michael F. (1992). "Geographical Information Science". *International Journal of Geographical Information Systems*, Vol. 6, No. 1, pp. 31-45. https://doi.org/10.1080/02693799208901893.
6. Haklay Mordechai, Weber Patrick (2008). "OpenStreetMap: User-Generated Street Maps". *IEEE Pervasive Computing*, Vol. 7, No. 4, pp. 12-18. https://doi.org/10.1109/MPRV.2008.80.
7. Harris Charles R. və b. (2020). "Array Programming with NumPy". *Nature*, Vol. 585, pp. 357-362. https://doi.org/10.1038/s41586-020-2649-2.
8. Huff David L. (1964). "Defining and Estimating a Trading Area". *Journal of Marketing*, Vol. 28, No. 3, pp. 34-38. https://doi.org/10.1177/002224296402800307.
9. Mammadov G., Mammadova S., Yusifova M., Sadigov R., Ahmedova G. (2025). "Agroecological fertility model and management of vineyard soils of the Lankaran-Astara economic region of Azerbaijan". *International Journal of Agriculture and Biology*, Vol. 34, Article 340104. https://doi.org/10.17957/IJAB/15.2339.
10. McKinney Wes (2010). "Data Structures for Statistical Computing in Python". In: van der Walt S., Millman J. (eds.). *Proceedings of the 9th Python in Science Conference*. Austin, Texas, pp. 56-61. https://conference.scipy.org/proceedings/scipy2010/mckinney.html (access date: 21.05.2026).
11. Mustafayev MG (2025). "Salts deposition and groundwater mineralization effect on soil salinization in the Shirvan Plain and Ganja-Gazakh Massif, Azerbaijan". *SABRAO Journal of Breeding and Genetics*, Vol. 57, No. 5, pp. 2147-2155. http://doi.org/10.54910/sabrao2025.57.5.35.

**12.** Sadigov RA (2025). “Geo-information Modeling and Spatial Analysis of Environmental and Infrastructure Parameters in the Shusha Region Using Satellite Imagery and GIS Technologies”. *International Conference on Current Problems in Engineering and Applied Sciences (ICCPEAS 2025)*, pp. 638-651.

**13.** Sadigov RA, Nazarova MK, Garayeva IE, Mammadova GI, Hasanova TA, Muhammad M (2025). “Modern Research of Using Alternative Energy Resources in Azerbaijan”. *International Journal of Advances in Applied Sciences*, Vol. 14, No. 3, pp. 907-915. http://doi.org/10.11591/ijaas.v14.i3.pp907-915.

**14.** Sinnott Roger W. (1984). “Virtues of the Haversine”. *Sky & Telescope*, Vol. 68, No. 2, p. 159.

**15.** Sultan Huseynova Maharram “Study of Modern Pasture Landscapes in Subtropical Zones of the Azerbaijan Republic”. *Environmental Pollution Journal*, Vol. 5, No. 2, pp. 185-192.

**16.** Vaswani Ashish və b. (2017). “Attention Is All You Need”. *Advances in Neural Information Processing Systems*, Vol. 30, pp. 59980-6008. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

## XÜLASƏ

Bu məqalədə Azərbaycanda biznes lokasiyalarının qiymətləndirilməsi üçün süni intellekt dəstəkli coğrafi qərar dəstəyi prototipi təqdim olunur. Tədqiqat 11 807 obyekt və 69 obyekt tipini əhatə edən GeoJSON datasetinə, Bakı rayonları üzrə demoqrafik göstəricilərə və Streamlit tətbiqində qurulmuş Python əsaslı məkan skorlaşdırma modelinə əsaslanır. Böyük dil modeli yalnız istifadəçinin təbii dildə yazdığı biznes ideyasını strukturlaşdırılmış niyyətə çevirmək üçün istifadə edilir. Yekun lokasiya tövsiyəsi isə Haversine məsafəsi, rəqib sıxlığı, dəstək obyektləri, tələb göstəricisi və demoqrafik üstünlük əsasında deterministik hesablama ilə formalaşdırılır. Nəticələr xəritə vizuallaşdırması və izah edilə bilən skor mexanizmi ilə təqdim olunur. Yanaşma kiçik və orta bizneslər üçün ilkin məkan seçimini daha ölçülə bilən, müqayisəli və yoxlanıla bilən qərar prosesinə çevirir.

**Açar sözlər:** *biznes lokasiyası, coğrafi informasiya sistemi, süni intellekt, Haversine məsafəsi, GeoJSON, lokasiya skoru, qərar dəstəyi*

**JEL kodları:** ***C63, C88, L81, R12, R32***