Enhancing Retail Location Decisions in City of Brno: An Application of Geospatial Analysis Tools

Oleksandr Turytsia[1] [0009-0002-1890-2451], Kristýna Zaklová[1] [0009-0009-5716-7446], Jiří Hynek[1] [0000-0002-7292-6094], Petr John[1] [0009-0000-8547-8510] and Tomáš Hruška[1] [0000-0001-5454-6592]

[1] Faculty of Information Technology, Brno University of Technology, Božetěchova 1/2, 612 66 Brno, Czech Republic
xturyt00@stud.fit.vutbr.cz
{zaklova,hynek,ijohn,hruska}@fit.vut.cz

Abstract. Location plays a key role in the success of a business. No amount of property features such as building, decorating, or price can overcome the negative impact of a poor location. A strategically positioned business not only reduces financial risks but also enhances the likelihood of achieving success. This work aims to develop a user-friendly information system to assist retailers in making informed location decisions. The system utilizes a well-known methodology based on Kernel Density Estimation, evaluating geo-demand and geo-competition, and the Analytical Hierarchy Process to determine a suitable location. The results are presented on the open datasets provided by the City of Brno, the second largest city in the Czech Republic. It allows users to choose between 84 types of business and display precalculated heatmaps of hot spots representing areas of a high range of commercial services. Then, the users are provided with the tool to evaluate their own locations of interest based on their own defined criteria. The results demonstrate the practical use of theoretical methodology with real data, evaluating its usability and performance aspects.

Keywords: analytical hierarchy process, decision support system, geo-visualization, kernel density estimation, location-based decision-making, open data, retail site decision process.

1 Introduction

What are the three most important factors in selling real estate? Location, location and location. This is applicable both in the private and public sector and also in the retail business [2, 7, 14]. Deciding where to locate business has always been a problem that people continuously tried to solve all over the world. Throughout time, a great majority of retailers would make a decision based on personal experience and instinct, regarding the process very much as an *art*. People would mainly use very subjective techniques. Some of them are no more than *hunches* based upon experience [16]. However, with the passage of time, it can be observed that the adoption of objective evaluation techniques in the field of retail location planning is increasing beyond employing simple *experience* to underpin store portfolio decision-making [26].

On the other side, classical location theories based on GIS techniques and statistical analysis make it possible to relate store location to distance from the center, population density and the location of competitors [25]. When relying on store-location models, it should be taken into account that these models are not always as comprehensive as they should be: they do not account for the impact of location-specific operating costs on location decisions; they do not account for the impact of a consumer's free time on their propensity to shop at a specific location; they do not include the impact of the internet and mail order, etc. [22] It is also necessary that such models do not become “black boxes” that only statisticians and domain specialists understand [36].

In the retail environment, businesses are surrounded by enormous amount of data and variables that can influence their success. Any entrepreneur embarking on the journey of starting a business faces a significant obstacle at the outset: the need for significant investment. Initial capital costs can be very challenging, from purchasing inventory to setting up infrastructure and hiring employees. However, customers rarely come immediately, and the initial investment doesn't start to pay off promptly.

As information systems evolved, research procedures became more sophisticated. This presented a challenge for retailers: without using location decision procedures to improve objectivity, they risked falling behind businesses that adopted such methodologies [16]. If retailers choose to do their own analysis, they are faced with large amounts of data, and browsing maps or websites to analyze competitors is time-consuming and with uncertain results. Retailers must carefully select and coordinate the tools to ensure they complement each other and provide a comprehensive view of the decision at hand. Otherwise, they risk making false decisions or mistakes.

The advantages of open data are realized by both the government and the private sector, with the main benefits including greater transparency, increased efficiency and effectiveness, and societal and economic contributions [35]. Such data can serve as a new resource for citizens, and its use holds the promise of social innovation if citizens realize the potential of this resource [21]. Since geospatial data plays a significant role in selecting a suitable location, relevant open datasets have been selected to confirm the possibility of using such data for new valuable applications.

We aim to build a system to aid retailers in making informed location decisions. The system can help with both possible scenarios when the retailer chooses a suitable retail site location—both in the case that the retailer has selected several suitable areas and needs to choose the right one from them; so if the retailer is not clear about the location and needs to find a suitable place for his business based on an own analysis. Therefore, the system is suitable both for identifying high-potential areas and performing competitive analysis and location evaluation. The chosen methodology outlined in [27] enables users to analyze multiple datasets, utilize geographical information systems features for site selection and input their preferences into the system, making the process flexible and suitable for any retailer who chooses to use it. Expected outputs include a generic model for geospatial data, which will be the input of a user-friendly analytical tool that is usable without the need to be a data analyst or advanced user of information technologies. This tool will offer a straightforward evaluation process for selecting a suitable location for a business. The results will be practically demonstrated using data from the city of Brno.

2 Retail Site Decision Process

The retail sector is currently in a phase of transformation due to various factors, including the rise in consumer mobility, the prevalence of e-commerce, shifting household demographics, and market saturation. With that, there were evolving business strategies to align with emerging trends. Since the late 1970s, there has been a noticeable decline in the growth of hypermarkets, with small and medium-sized supermarkets gaining preference [20]. This shift in consumer behavior has forced small retailers to adopt more strategic approaches when selecting store locations, determining store size, and offering services.

The problem of retail site decision-making is a complex process composed of several steps, beginning with an analysis of the retail area (e.g., distribution of potential customers, available markets, etc.), followed by specification of required criteria (e.g., accessibility, sales floor, etc.), and determination and evaluation of suitable location candidates. Roig-Tierno et al. [27] present a methodology which combines the application of Geographical Information Systems (GIS) with the Kernel Density Estimation (KDE) method, followed by the determination of possible location using the Analytical Hierarchy Process (AHP), shown in Fig. 1.

The concept of the process is based on spatial dispersion: geo-demand (the location of potential customers) and geo-competition (the location of business competitors) [4], which are used to calculate KDE. In the next step, the decision-maker must select the areas of interest and provide attributes for each potential location, which he can then compare against each other using a scale. Once it is done, AHP is applied to evaluate all the attributes on each site and output locations with their rating. The location with the highest rating is considered to be the most desired [27].

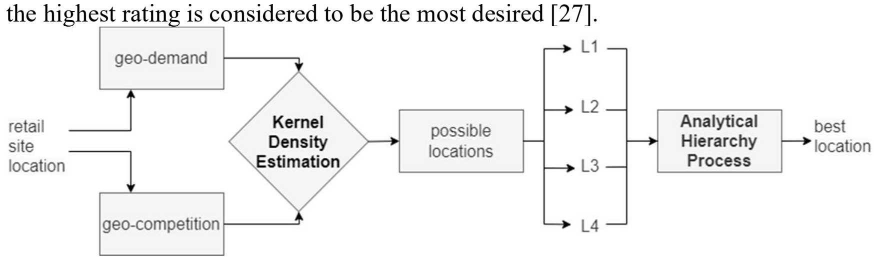

Fig. 1. The entire retail site location decision process consists of several consecutive steps. The letters L with a number indicate specific locations on the map selected by the decision-maker. This concept was taken from the publication [27].

As [8, 11, 19] shows, the process can be adjusted to specific needs, or other methodologies can be used. This section aims to conduct a concise analysis of the algorithms and methodologies that can be adopted for the presented process.

2.1 Geo-demand

According to [27], *geo-demand* can be defined as the location of potential customers who purchase a product or service in a specific market. Individuals who live in the area can be viewed as potential customers. This is due to the possibility that individuals might express interest in various markets without certainty regarding their preferences and characteristics (e.g., age, gender, income level, etc.). The data for geo-demand can be acquired from the local city database. In practice, it is a set of points projected onto a map (Fig. 2).

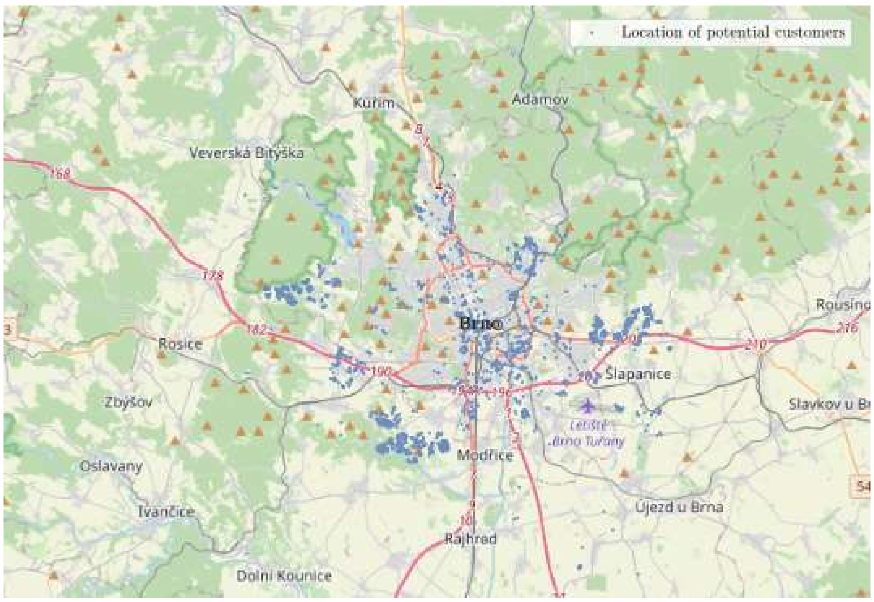

Fig. 2. A simple example of geo-demand in Brno. Each blue point represents the location of potential customers. It contains the number of people living at a specific address.

2.2 Geo-competition

Geo-competition is the location of a business's competitors and the delineation of their trade areas in a particular market. Trade area can be defined as the geographic area in which a retailer attracts customers [4, 27]. It is a more complex task than geo-demand identification. Theoretical methods for defining trading areas, which have grown significantly in the last decade, involve developing conceptual frameworks or models that explain the spatial interactions and patterns associated with trade or business activities. These methods are based on theoretical principles about how certain factors influence the formation of trading areas [20].

There are two main types of these models:

- **Gravity Models—**theoretical frameworks borrowed from physics. They are derived from the laws of Newtonian physics, based on the balance between the store attractivity and the distance to the potential customers [20].
- **Probabilistic Models—**are based on the likelihood of customers visiting a certain location within a specific geographical region [18].

Huff Model is an example of a probabilistic model presented by David L. Huff [18]:

$$P_{ij} = \frac{\frac{S_j}{T_{ij}^{\lambda}}}{\sum_{j=1}^{n} \frac{S_j}{T_{ij}^{\lambda}}} \tag{1}$$

where:

- P_{ij}: the probability of a consumer at a given point of origin i traveling to a particular shopping center j;
- S_j: the size of shopping center j (measured in terms of the square footage of selling area);
- T_{ij}: the travel time involved in getting from a consumer's travel base i to a given shopping center j;
- λ: a parameter to be estimated empirically to reflect the effect of travel time on various shopping trips.

According to [29], the shopper's behavior is affected by the image, which comprises elements such as brand reputation, values, personality, and associations. These factors can be represented using parameters like the *distance-decay* factor in the original model [15]. Another possibility is to use data from social media, thanks to which an overview of consumer behavior can be obtained. However, it's crucial to acknowledge that social media data has limitations—it only captures the activities users share online and doesn't fully represent all their real-world actions [34].

2.3 Kernel Density Estimation

The KDE method can be applied to match the information from geo-demand and geo-competition analysis to obtain the area where the high population has a poor range of commercial services [27]. In essence, the objective of KDE is to compute the density of points within a specified area based on the distances between the points.

According to [27], a pixel was adopted as a unit of analysis. A pixel is a square on a digital map that represents a specific area and is assigned a value linked to the features within that space. For each square on the map, a circular area is created using the square's center as the circle's center. The data points within this circle can have different weights to these, considering their distance from the center of the square. Simply put, points closer to the center have more influence, while those farther away carry less weight (Fig. 3). This concept can be based on [27] expressed as follows:

$$L_j = \sum_{i \in C_j} \frac{3}{\pi r^2}\left(1 - \frac{d_{ij}^2}{r^2}\right)^2 \tag{2}$$

where:

- L_j: the estimated density of a pixel;
- d_{ij}: the distance between points i and j;
- r: the width of the window or search radius, which determines the degree of smoothing;
- $C_j = \{i|d_{ij} < r\}$: the set which consists of the i points whose distances from the centroid of the pixel are less than the established radius of the circle.

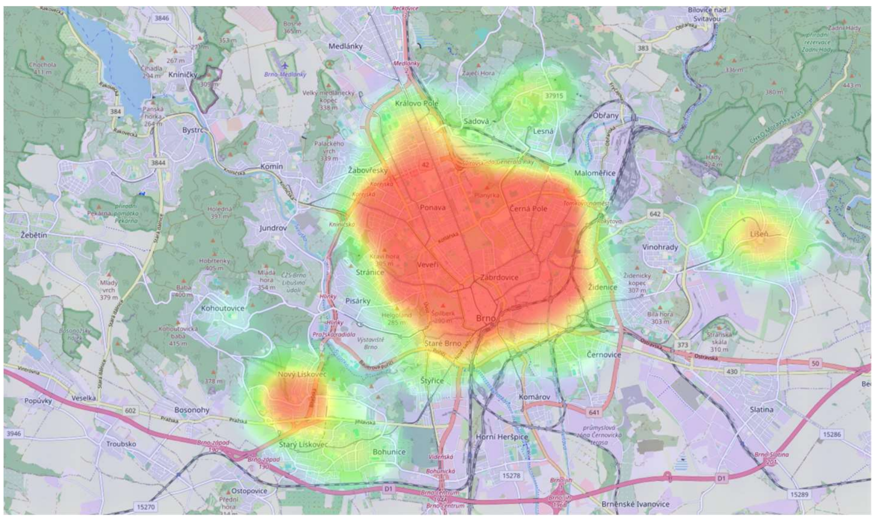

Fig. 3. An Example of KDE on a map. Hot spots represent areas with potential customers and a high range of commercial services. This figure was made using a dataset with population and the first 35 candy stores from the dataset with retail outlets of Brno.

2.4 Determining Optimal Locations

Businesses often employ advanced mapping techniques integrating demographic data and service coverage to identify optimal retail locations in areas with a lack of services [27]. A crucial aspect of this process is to create heatmaps using KDE, which provides a visual representation of customer density and service gaps across different regions (Fig. 3). Once the heatmap is estimated, it becomes possible to pinpoint regions with customers and poor range of services (Fig. 4).

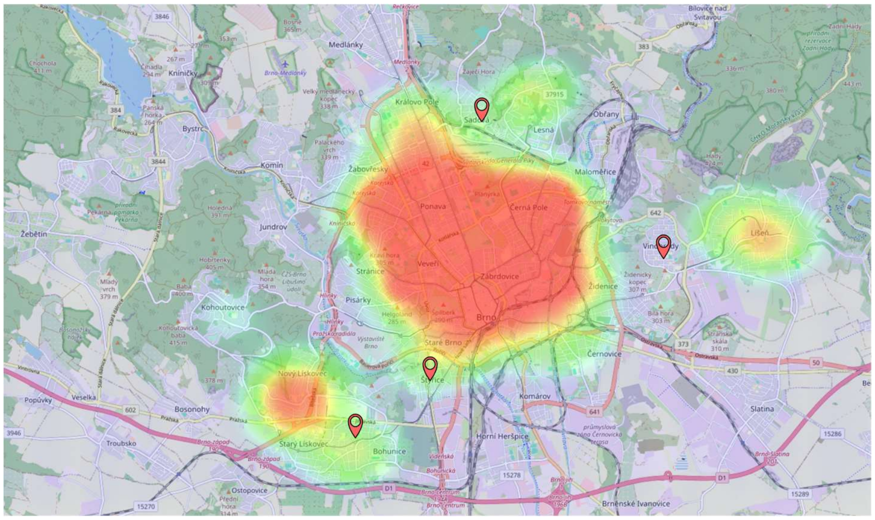

Fig. 4. An example of selecting possible locations after the KDE using the same estimated kernel density as in Fig. 3. The markers represent possible locations specified by the users. They could also be mined from rental services that offer available commercial space.

2.5 Multi-Criteria Decision Analysis

Multi-Criteria Decision Analysis (MCDA) is a systematic approach employed in decision-making processes involving evaluating multiple alternatives. This approach addresses the complexity of decision problems by considering various criteria simultaneously, providing a structured framework for assessing and comparing potential options [10]. Multi-criteria methods usually have in common that most decisions can be improved by decomposing the overall evaluation of alternatives into evaluations on several criteria relevant to the problem [10]. To help retailers evaluate multiple criteria, there are a variety of methods [37]—such as PROMETHEE, ELECTRE, TOPSIS, **AHP** or hybrid methodologies like fuzzy AHP [1] etc.

The PROMETHEE (*Preference Ranking Organization METHod for Enrichment Evaluations*) methods [5] belong to the family of the outranking methods. The essence of these methods is constructing an outranking relation on a finite set of actions and exploiting this relation to answer a multicriteria problem, considering each action's leaving and entering flow in the valued outranking graph.

The ELECTRE (*ELimination Et Choix Traduisant la REalité*) method [13, 32] works based on alternatives' comparison considering individual criteria and using important coefficients. Preferences are modeled by using binary outranking relations. In this method, it is impossible to offset a very bad value on a criterion with good ones on other criteria.

TOPSIS (*Technique for Order Preference by Similarity to Ideal Solution*) [12, 23] is a numerical compensatory method that allows the compromise between different criteria, where a bad result in one criterion can be compensated by a good result in another. The TOPSIS method assumes that each criterion has a monotonically increasing or decreasing preference.

AHP (*Analytical Hierarchy Process*) is a method which evaluates all the attributes on each site and outputs locations with their relative rating. The location with the greatest value of the rating is considered to be the most desired. This method takes advantage of that by structuring decision problem into a comprehensive hierarchy which consists of several levels [32, 33]. The first level of the hierarchy is a goal of the process; sub-levels are constructed of criteria that can affect the choice. The last level of the hierarchy contains the alternatives, which are assessed at the end of the process.

An important task of the selection process is to select suitable criteria. There are typical selection criteria and factors for the selection of retail site location that are used by different researchers [3, 6, 11, 27]. The main criteria (e.g. distance, traffic congestion, market attractiveness) and relevant sub-criteria are always set (e.g. distance, it can be the distance to buffets, distance to restaurants, etc.).

2.6 **Visualization of Geo-Spatial Data**

In today's rapidly advancing technological world, the integration of Geographical Information Systems (GIS) has become crucial in various fields, including retail location decision [30]. GIS offers a powerful toolset for capturing, analyzing, and visualizing spatial data, providing valuable insights that aid decision-making processes. GIS applications may include cartographic data, photographic data, digital data, or data in spreadsheets. A widely used format for representing geospatial data is GeoJSON[1], which provides a simple and lightweight structure for encoding different types of geometry.

Digitalizing real-world geospatial data is essential for working with and storing it on a computer. The utilization of basic geometric shapes is a common method to describe objects in the real world because they are suitable to be stored in database systems. These data types are commonly referred to as *spatial data types*, covering categories like point, line, and region. There are various web-based libraries (e.g. Leaflet, MapboxGL, etc.) which help to create a custom geospatial visualization, or frameworks (e.g., Geovisto) which provide ready-to-use thematic map templates (e.g. heatmap, dot map, etc.) [17]. Additionally, more complex types, such as partitions and graphs (networks), can be included [28, 31]. Spatial data types serve as a foundational abstraction, enabling modeling the geometric structure, relationships, properties, and operations with objects in space [28].

The current state of open geospatial data was reviewed by researchers in the 2020 survey [9], who concluded that the adoption of open data is also growing in the geospatial domain and that open geospatial data is likely to have an even deeper impact in the future. Source [24] focused on experimental investigations conducted on more than 160,000 geospatial datasets from six national and international open government data portals. The results showed a significant underutilization of geospatial datasets, a generally poor metadata quality, and a weak correlation between the use and quality of the metadata.

---

[1] https://geojson.org/

3 Problem Analysis

As described in Section 2, retail site decision-making is a complex process that demands a nuanced understanding of diverse methodologies. Such a requirement is not expected from the non-academic general public (such as start-up entrepreneurs, small business owners or local retailers) whose focus is instead on their domain of business. Our goal is to integrate open data sources provided by municipalities with a system providing user-friendly guidance through the steps of the process so that the user can use a suitable tool for every step of the process to evaluate the selected areas according to their own preferred criteria.

Since the aim is to provide the tool to non-professional individuals, the requirements of simplicity and usability of the user interface need to be met. Geospatial data visualizations must be easy to understand, and the entire process of evaluating the selected sites should be intuitive. The whole concept needs to be validated using a specific municipality and real data. Based on our extensive cooperation with the city of Brno, we decided to incorporate the municipality's open datasets as a pilot study for the present undertaking.

The city of Brno is the second largest city in the Czech Republic in terms of area and population. Currently, it has almost 400 thousand inhabitants. The City of Brno is aware of the growing importance of the digital world, which is why the Data department of Brno City Municipality employees have long made efforts to provide their data publicly. For this purpose, the city's data portal[2] was launched in 2018. Currently, the portal contains 376 datasets, of which 205 are published under some version of the CC BY license. Portal also offers a number of interactive overviews (Brno in numbers, interactive dashboards, etc.).

For the purpose of this work, two relevant datasets[3] were used. Both datasets follow the same initial data model based on GeoJSON standard format[4] (Listing 1) composed of the list of points. Based on the dataset, different properties are related to the points:

1. The **dataset with the number of persons living at addresses** has the following information in the *properties* attribute: number of people, house number, reference number and symbol and street name. Coordinates from this dataset represent a potential customer's location, which is required to calculate the distance between a customer and a competitor.
2. The dataset **Brno Retail Research** contains information about all the business outlets in Brno, such as area, type and many other less relevant attributes. In the *properties* attribute are essential: type of service (category of a business) and size of retail outlet (area of the outlet).

[2] https://datahub.brno.cz/

[3] Number of people living at the addresses (https://arcg.is/1Lfbzb0) and Brno Retail Research (https://arcg.is/0CaaCS)

[4] RFC specification: https://datatracker.ietf.org/doc/html/rfc7946

```
{
   "features": [
   {
      "type": "Feature",
      "properties": { ... },
      "geometry": {
        "type": "Point",
        "coordinates": [ 16.58, 49.17 ]
      }
    }, ... ]
}
```

Listing 1. The datasets provided by the city of Brno comprise the list of *points* in GeoJSON format. Every point is composed of custom properties related to the dataset.

4 Proposed Solution

The key tasks were to design the architecture of the proposed system, which handles input data preprocessing, implements the retail site decision process and provides a user interface to the end-users in order to perform the process from the users' point of view.

4.1 System Architecture

The system was developed as a full-stack web application with client-server architecture, which is shown in more detail in Fig. 5. The application's source code is publicly available[5], and the tool for the city of Brno has been deployed under a public address[6]. Now, the deployment directly to the server of the city of Brno is being negotiated.

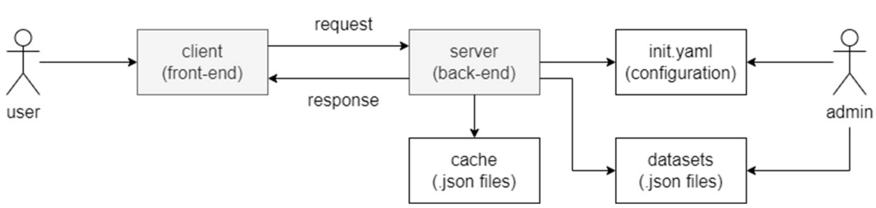

Fig. 5. A diagram of the system architecture. Target users include users who interact with the client part of the system and the administrator who defines the target municipality's initial configuration and input datasets.

The back-end side of the system was implemented using Python and the FastAPI library[7]. It runs as a standalone REST API application within the overall system architecture. It can be configured with various regions or datasets, including customer

[5] https://github.com/aturytsia/retail-site-location-decision-system
[6] https://dexter.fit.vutbr.cz/retail-site-dss
[7] https://fastapi.tiangolo.com/

data and competitor information, via a configuration file *init.yaml* in the root folder of the server[8]. The KDE is calculated from the input data while identifying areas with high competition, which is computationally demanding. For this reason, before the server startup, the initial task of the server is to read all the datasets, perform all the calculations and save the result in a *cache*. Specific calculations on the server part of the application will be detailed in section 4.3.

The front-end side is implemented using Typescript, React[9] and Leaflet library[10]. It allows the users to interact with spatial data. It has a dynamic map displaying various data layers and components presenting information relevant to each stage of the decision-making process. The specific use case from the user's point of view on the client parts of the application will be detailed in section 4.4. In addition to the process of sites evaluation, this part of the application offers the possibility to work with layers. This component is accessible during any step of the process. It can provide extra information for the users when selecting possible locations, and it allows users to overlay different datasets onto a single map, creating informative layers that provide a comprehensive view of relevant information, similar to GIS data layering. These layers include high-demand areas, grid layers, and customer and competitor locations.

4.2 Input Data Preprocessing

To determine geo-demand and geo-competition, it is necessary to use the geospatial data of a given municipality. The data in the GeoJSON format from the city of Brno were used for the practical demonstration, as introduced in Section 3. However, for different cities, the data format can vary. Hence, it was necessary to design a general internal data representation to unify the input data from different sources (Listing 2).

Both input datasets of the city of Brno were to be preprocessed with a Python script into the proposed format. In addition, for optimization purposes, the **Brno Retail Research** dataset was split into 84 datasets depending on the business category.

```
[
     [latitude_1, longtitude_1, data_1],
     [latitude_2, longtitude_2, data_2],
     ...,
     [latitude_n, longtitude_n, data_n]
]
```

Listing 2. Internal data representation is made up of triples—the latitude and longitude position of a point and associated information or data related to the corresponding coordinates (e.g. the number of customers or the size of a competitor).

[8] https://github.com/aturytsia/retail-site-location-decision-system/blob/main/server/init.yaml
[9] https://react.dev/
[10] https://leafletjs.com/

4.3 Calculations on the Server Side

On the application's server side, two essential calculations take place—high competitive areas calculation using the input datasets and the KDE method and the AHP evaluation using user-defined attributes, weights and scores.

A function for high competitive areas calculation is called whenever the server is launched in order to start the evaluation. It is important to mention that this function requires a graph of the city of Brno provided by the *osmnx*[11] library. The graph is installed based on the specified region in the configuration and then used in the function to find distances between customers and competitors. A diagram of this process is shown in Fig. 6. The inputs are datasets with information about customers and competitors and the distance decay factor for the Huff model. This factor is used to model the phenomenon of distance, which is perceived as a non-linear deterrent to motion and is related to the type of competitors[12].

The AHP procedure is invoked during the final stage of the process to evaluate potential locations, considering both location attributes and attribute importance defined by the user. The result is a numerical value between 0 and 1 for each selected location. In addition to the AHP results, a consistency ratio[13] that indicates the consistency between pairwise comparisons is also calculated.

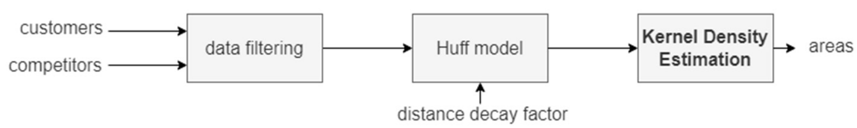

Fig. 6. Partial steps of the process of highly competitive areas calculation.

The process of highly competitive areas calculation includes these steps:

1. **Data Filtering:** if needed, unnecessary records are filtered out (e.g. those not relevant to the specified field).
2. **Huff model:** it was introduced in more detail in Section 2.2. The algorithm for the calculation is described in Listing 3; the output of this step is an array of averaged probabilities for all customers.
3. **KDE:** it was introduced in more detail in Section 2.3. The essence of the step is creating a distribution of probabilities across the map. The results are the basis for the heatmap layer.

---

[11] https://osmnx.readthedocs.io

[12] E.g. people only go shopping for food in nearby shops, while they are willing to travel to more distant shops to buy furniture: https://pro.arcgis.com/en/pro-app/latest/tool-reference/business-analyst/understanding-huff-model.htm

[13] https://www.spicelogic.com/docs/ahpsoftware/intro/ahp-consistency-ratio-transitivity-rule-388

```
for c in competitors:
  dest_node = get_nearest_node(square_key, c.x, c.y)
  for m in customers:
    current_node = get_nearest_node(square_key, m.x, m.y)
    distance = get_distance_to_node(dest_node, current_node)
    travel_time = distance / AVERAGE_WALKING_SPEED
    p = get_probability(area, time, distance_decay) / p_sum
```

Listing 3. Calculation of the Huff model. The position of a competitor ***c*** or customer ***m*** on the map is calculated as the center of the cell (i.e. centroid) in which the entity is located. The resulting 2D array is then converted into a table where rows correspond to customers and columns correspond to competitors. For each competitor, the average probability ***p*** of customer visits across all stores on the map is computed.

Rather than computing the distance individually between each customer and competitor in Huff model calculation, a grid-based approach will be used (Fig. 7), covering the entire map with cells measuring 500 meters each. Assigning competitors and customers to their respective cells simplified the process of calculating the distance between cells. Leveraging this grid structure allowed for the efficient reuse of distances, particularly beneficial in scenarios where cells contained multiple competitors and customers.

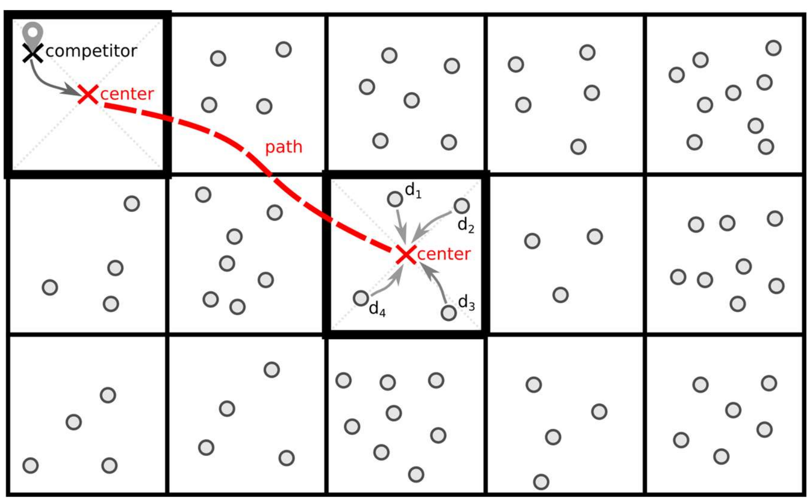

Fig. 7. The map is divided into a grid of cells. Every competitor and every customer (d_i) is mapped in the center of a corresponding grid cell. Then, paths between competitors and customers are considered as the paths between the corresponding centroids. Such optimization causes tolerable inaccuracies in the results on the one hand but significant performance improvement on the other.

Despite efforts to introduce dynamic programming and optimisation techniques, the function's performance in identifying areas with high competition was not good enough. For this reason, the results of this function (i.e. precomputed KDE) are stored in a cache that must be cleared if new calculations are needed. Before the server startup, the initial task of the server is to read all the datasets, perform all the calculations and save the result.

4.4 User Process of Sites Evaluation

As mentioned in Section 3, concerning the target users, a step-by-step workflow must be offered to assist the users in conducting location evaluations. The process consists of chronological steps (Fig. 8), which the users will manage. During the process, they will be asked to provide additional input representing their own criteria that reflect the specific needs of the users.

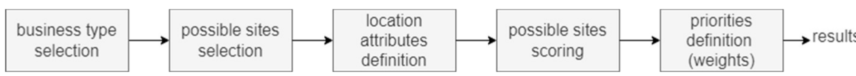

Fig. 8. Partial steps of the process of evaluating selected locations from the user's perspective.

These steps include:

1. **Business Type Selection:** the user selects the type of business they are interested in (e.g. books, confectionery or drugstores).
2. **Possible Sites Selection:** the system displays a heatmap for the selected business type on a map of the city of Brno. The aim is to facilitate the user's selection of potential locations. The heatmap is created based on the source data of the examined city, thanks to which it is possible to identify geo-demand and geo-competition. Once these concepts are identified, the system can apply kernel density estimation. Combining these concepts makes it possible to identify areas with a lack of competitors relative to customers. The users mark the preferred locations on the map with the help of the heatmap layer.
3. **Location Attributes Definition:** once the potential sites are selected, the user must define at least three location attributes. These can be accessibility, parking, sales floor or anything else that can be described using quantitative or qualitative information.
4. **Possible Sites Scoring:** each defined attribute must be scored for each location. This process is entirely up to the user, and to perform the evaluation correctly, the user must have a deeper knowledge of the selected sites. The rating can be done at two levels—using a five star scale or numerical attributes, specifying a value and a maximum value.
5. **Priorities Definition:** the defined weights are entered into the AHP method. They can be set on a scale of 1 to 9, with a default value of 1 for each attribute.

The result is a numerical value for each selected point on the map of Brno. The highest value represents the best location estimated based on user input in the previous steps.

5 Results

The results demonstrated practically the theoretical methodology in a specific municipality. The people considering locating their business in the city of Brno have received a tool that should help them choose the right location. In addition, the information system was developed with an emphasis on possible applicability in other cities as well. This would include getting the necessary datasets, transforming them into an internal data model, and configuring the server properly. So it shouldn't be difficult to bring this solution to other cities.

The reliability of the application was tested by unit tests on both server and client parts of the application. Within the application in the city of Brno, it is possible to evaluate the usability of the solution and performance aspects. The client part contains two main components in accordance with the proposed solution—the first represents the user process of sites evaluation (part of the process in Fig. 9); the second is a dynamic map with various data layers (Fig. 10). User friendliness, clarity and usability were tested with a selected sample of users from the city of Brno. The testing revealed minor deficiencies in comprehending the procedural steps of evaluating sites. These shortcomings were eliminated by adding accompanying texts and user help.

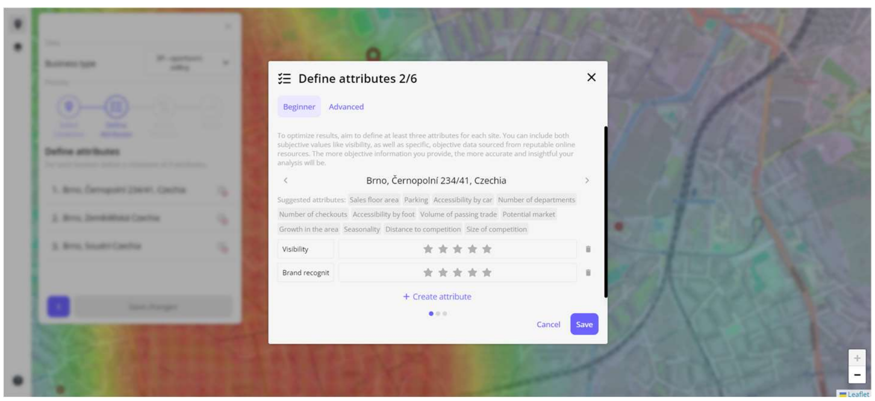

Fig. 9. Screenshot of the result application with the user interface, which shows the part of the process dealing with the definition of location attributes and scoring of selected sites of interest. In this case, the rating scale is the default, i.e. five stars.

In the context of this project, performance testing is essential due to the complexity of the function for calculating geo-competitors, i.e. areas with high competition for the provided dataset. The time taken for these calculations can vary significantly based on the size of the datasets provided because of the $O(n^2)$ time complexity of the algorithm. For this reason, the heatmaps are precalculated and cached for datasets of Brno, and it is not necessary to re-run this during the runtime of the server. For Brno, the dataset is represented by 29 020 data records of customers and 8 335 data records of competitors divided into 84 types of business. The evaluation of all the provided datasets took a total of 1 971.31 seconds on an Apple M3 GPU (14-core/36GB/30-core). It's worth mentioning that the results may differ depending on the number of threads available in the system where the evaluation was initiated. Optimization should be the focus of future work.

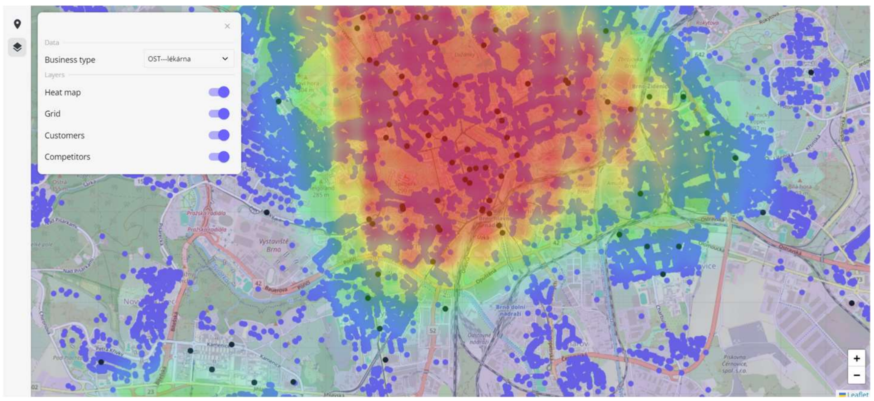

Fig. 10. Image of the resulting application with layers on the Brno city map for the dataset of pharmacies. High-demand areas where commercial service is poor and population density is high are represented by a heatmap (Section 2.3); the grid shows the division of the city of Brno into cells; customers are shown as purple dots and competitors as black dots. All these layers can be displayed or hidden independently.

6 Discussion

The development of such a system represents a significant step forward in empowering retailers to make more informed location decisions. Another target user may be the management of the city of Brno, as the tool can identify which services are missing in the chosen districts of the city. Using the outputs of the system, they can support the provision of missing services to the citizens.

The main advantage of this work is the simplicity of the system—analytical tools are typically quite complex, and not everyone can use them. When creating these results, the emphasis was placed on the fact that even a less technically skilled user would be able to use them, for whom the tool can facilitate site selection and, as a result, save costs. The application of the results has only been carried out in one city, but nothing prevents their use in other municipalities.

The unavailability of the source data for the created tool may cause a potential problem. Although the trend towards open data is growing, municipalities do not always provide the relevant data. There will always be some dependence on the willingness of municipalities to make their data public. It is, however, worth saying that this is an activity in their interest—a well-chosen location is a key factor for a successful business.

Due to the lack of standardization of the open datasets format, it is also necessary to consider that pre-processing of the data will be required for possible integration into the tool. Also, it is important to remember that despite the huge advances in site assessment techniques and tools, there is no substitute for the field visit and the observations across various spatial scales and times of day [36].

Further research could focus on a deeper analysis of potentially useful datasets of the city of Brno and integrate these datasets with the site scoring step of the process. Currently, users need to specify their own criteria for the AHP method. The fulfillment of this step might be a problem since users might not be able to evaluate the candidate sites objectively. In the future, the datasets of the Brno public transport system or the historical traffic data from the WAZE navigation available in the Brno data catalog might be used.[14]

7 Conclusions

This document describes the retail site location selection problem and how open data can be helpful in dealing with it. This paper uses a methodology combining GIS and multi-criteria decision models [27] and demonstrates it in a practical way to make it easier for retailers to select a suitable location for their business. For this purpose, three key concepts were developed: the model for geospatial data, the analytical tool for use by the general public and the evaluation process for selecting a suitable location. The proposed solution has been demonstrated on open data from the city of Brno and the results show that both applicability to other cities and refinement of the quality of the calculations by adding new relevant datasets is possible.

Acknowledgements. This work was supported by project Smart information technology for a resilient society, FIT-S-23-8209, funded by Brno University of Technology.

References

1. Al Garni, H. Z., Awasthi, A.: A fuzzy AHP and GIS-based approach to prioritize utility-scale solar PV sites in Saudi Arabia. In 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1244-1249 (2017).
2. Anderson, S. P., Goeree, J. K., & Ramer, R.: Location, location, location. Journal of economic theory, 77(1), 102-127 (1997).
3. Baviera-Puig, A., Buitrago-Vera, J., Escriba-Perez, C.: Geomarketing models in supermarket location strategies. Journal of Business Economics and Management, 17(6), 1205-1221 (2016).
4. Baviera-Puig, A., Buitrago-Vera, J., Mas-Verdú, F.: Trade areas and knowledge-intensive services: the case of a technology centre. Management decision, 50(8), 1412-1424 (2012).
5. Brans, J. P., Vincke, P., Mareschal, B.: How to select and how to rank projects: The PROMETHEE method. European journal of operational research, 24(2), 228-238 (1986).
6. Burnaz, S., & Topcu, Y. I.: A multiple-criteria decision-making approach for the evaluation of retail location. Journal of Multi-Criteria Decision Analysis, 14(1-3), 67-76 (2006).
7. Celik Turkoglu, D., Erol Genevois, M.: A comparative survey of service facility location problems. Annals of Operations Research, 292, 399-468 (2020).
8. Clarke, I., Rowley, J.: A case for spatial decision-support systems in retail location planning. International Journal of Retail & Distribution Management, 23(3), 4-10 (1995).

[14] https://datahub.brno.cz/

9. Coetzee, S., Ivánová, I., Mitasova, H., Brovelli, M. A.: Open geospatial software and data: A review of the current state and a perspective into the future. ISPRS International Journal of Geo-Information, 9(2), 90 (2020).
10. Durbach, I. N., Stewart, T. J.: Modeling uncertainty in multi-criteria decision analysis. European journal of operational research, 223(1), 1-14 (2012).
11. Erbıyık, H., Özcan, S., Karaboğa, K.: Retail store location selection problem with multiple analytical hierarchy process of decision making an application in Turkey. Procedia-Social and Behavioral Sciences, 58, 1405-1414 (2012).
12. Erdin, C., Akbaş, H. E. A comparative analysis of fuzzy TOPSIS and geographic information systems (GIS) for the location selection of shopping malls: a case study from Turkey. Sustainability, 11(14), 3837 (2019).
13. Figueira, J. R., Mousseau, V., Roy, B: ELECTRE methods. Multiple criteria decision analysis: State of the art surveys, 155-185 (2016).
14. Formánek, T., Sokol, O.: Location effects: Geo-spatial and socio-demographic determinants of sales dynamics in brick-and-mortar retail stores. Journal of Retailing and Consumer Services, 66, 102902 (2022).
15. France, Stephen L., and Sanjoy Ghose. Marketing analytics: Methods, practice, implementation, and links to other fields. *Expert Systems with Applications* 119: 456-475. (2019).
16. Hernandez, T., Bennison, D.: The art and science of retail location decisions. International Journal of Retail & Distribution Management, 28(8), 357-367 (2000).
17. Hynek, J., Rusňák, V.: Towards Interactive Geovisualization Authoring Toolkit for Industry Use Cases. International Joint Conference on Computer Vision, Imaging and Computer Graphics. Cham: Springer International Publishing (2021).
18. Huff, D. L.: Defining and estimating a trading area. Journal of marketing, 28(3), 34-38 (1964).
19. Lin, G., Chen, X., Liang, Y.: The location of retail stores and street centrality in Guangzhou, China. Applied geography, 100, 12-20 (2018).
20. Mendes, A. B., Themido, I. H.: Multi-outlet retail site location assessment. International Transactions in operational research, 11(1), 1-18 (2004).
21. Morelli, N., Mulder, I., Concilio, G., Pedersen, J. S., Jaskiewicz, T., de Götzen, A., Arguillar, M.: Open Data as a New Commons. Empowering citizens to make meaningful use of a new resource. In: Internet Science: 4th International Conference, Proceedings 4, pp. 212-221. Springer, Greece (2017).
22. Nwogugu, M.: Site selection in the US retailing industry. Applied mathematics and computation, 182(2), 1725-1734 (2006).
23. Pavić, Z., & Novoselac, V.: Notes on TOPSIS method. International Journal of Research in Engineering and Science, 1(2), 5-12 (2013).
24. Quarati, A., De Martino, M., Rosim, S.: Geospatial open data usage and metadata quality. ISPRS international journal of geo-information, 10(1), 30 (2021).
25. Reigadinha, T., Godinho, P., Dias, J.: Portuguese food retailers–Exploring three classic theories of retail location. Journal of Retailing and Consumer Services, 34, 102-116 (2017).
26. Reynolds, J., Wood, S:. Location decision making in retail firms: evolution and challenge. International Journal of Retail & Distribution Management, 38(11/12), 828-845 (2010).
27. Roig-Tierno, N., Baviera-Puig, A., Buitrago-Vera, J., Mas-Verdu, F.: The retail site location decision process using GIS and the analytical hierarchy process. Applied Geography, 40, 191-198 (2013).
28. Schneider, M.: Spatial data types: Conceptual foundation for the design and implementation of spatial Database systems and GIS. In Proceedings of 6th International Symposium on Spatial Databases (1999).
29. Stanley, T. J., Sewall, M. A.: Image inputs to a probabilistic model: predicting retail potential. Journal of Marketing, 40(3), 48-53 (1976).

30. Suárez-Vega, R., Santos-Peñate, D. R., Dorta-González, P.: Location models and GIS tools for retail site location. Applied Geography, 35(1-2), 12-22 (2012).
31. Sümer, S. I., Sümer, E., Atasever, H.: Promoting development through a geographic information system-based Lodging Property Query System (LPQS) for Antalya, Turkey. Information Development, 32(4), 1055-1067 (2016).
32. Taherdoost, H., Madanchian, M.: A Comprehensive Overview of the ELECTRE Method in Multi-Criteria Decision-Making. Taherdoost, H., Madanchian, M, 5-16 (2023).
33. Tur-Porcar, A., Roig-Tierno, N., Llorca Mestre, A.: Factors affecting entrepreneurship and business sustainability. Sustainability, 10(2), 452 (2018).
34. Wang, Y., Jiang, W., Liu, S., Ye, X., & Wang, T.: Evaluating trade areas using social media data with a calibrated huff model. ISPRS International Journal of Geo-Information, 5(7), 112 (2016).
35. Welle Donker, F., Van Loenen, B., Bregt, A. K.: Open data and beyond. ISPRS International Journal of Geo-Information, 5(4), 48 (2016).
36. Wood, S., & Tasker, A.: The importance of context in store forecasting: The site visit in retail location decision-making. Journal of Targeting, Measurement and Analysis for Marketing, 16, 139-155 (2008).
37. Yap, J. Y. L., Ho, C. C., Ting, C. Y.: A systematic review of the applications of multi-criteria decision-making methods in site selection problems. Built environment project and asset management, 9(4), 548-563 (2019).