Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 266 (2025) 802–809

Procedia Computer Science

www.elsevier.com/locate/procedia

12th International Conference on Information Technology and Quantitative Management (ITQM 2025)

Profiling Healthy Neighborhoods Through Retail Food Spatial Analysis: The Case of Madrid

Ramón Alberto Carrasco[a*], Ziwei Shu[a*], Christiam Méndez-Lazarte[b], Mar Souto Romero[c]

[a]*Department of Marketing, Faculty of Statistics, Complutense University of Madrid, Madrid 28040, Spain*
[b]*Universidad de Lima, Av. Javier Prado Este 4600, Santiago de Surco, Lima 15023, Perú*
[c]*Department of Business Economics, Rey Juan Carlos University, Madrid 28032, Spain*

Abstract

Creating healthy neighborhoods is essential for enhancing residents' quality of life, with the local retail food environment being a key factor. Madrid's neighborhoods vary significantly regarding access to nutritious food options, potentially linked to their primary function as residential or tourist areas. This study proposes and applies a structured methodology, adapted from the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, to profile the city's neighborhoods using a modified Retail Food Environment Index (mRFEI) derived from OpenStreetMap (OSM) data. OSM, an open-source platform providing freely accessible geographic information, was used to identify and classify food outlets as “Healthy” or “Unhealthy” based on specific shop and amenity tags selected via expert knowledge. The mRFEI, calculated as the ratio of healthy to unhealthy outlets per neighborhood, was analyzed spatially using geospatial techniques. Results, visualized via interactive maps differentiating mRFEI scores below and above 100, reveal significant heterogeneity across Madrid, often indicating a lower relative presence of “Healthy”-classified outlets in central areas compared to several peripheral ones. By mapping and analyzing the distribution of food retailers, this study generates valuable insights to support informed policymaking aimed at promoting equitable access to healthy food options and contributing to health-focused urban development in Madrid. While demonstrating the utility of open data for large-scale assessment, this study acknowledges limitations inherent in OSM data and index simplification, suggesting avenues for future validation and refinement.

Peer-review under responsibility of the scientific committee of the 12th International Conference on Information Technology and Quantitative Management

Keywords: Decision-Making; modified Retail Food Environment Index; OpenStreetMap; Healthy Neighborhoods; Madrid

* Corresponding author.
E-mail address: ramoncar@ucm.es; ziweishu@ucm.es

1877-0509 © 2025 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 12th International Conference on Information Technology and Quantitative Management
10.1016/j.procs.2025.08.100

1. Introduction

Assessing the health characteristics of urban environments is crucial for effective public health policy and urban planning, particularly in understanding the retail food environment—the physical and built settings where food is accessible to the population [1, 2]. The local food environment, specifically the type and distribution of food retail outlets, significantly influences dietary patterns and health outcomes [3]. However, characterizing these environments consistently and scalable across large urban areas remains challenging, often limited by data availability and methodological complexities [4]. Indices such as the Retail Food Environment Index (RFEI) provide quantitative measures for assessing and visualizing the broader food environment, but their effectiveness depends on the availability of reliable data and clearly defined implementation frameworks [5].

Despite the growing interest in urban food environments, existing approaches often lack methodological standardization, reproducibility, and integration with reliable data mining workflows [6]. This study addresses these gaps by proposing a structured methodological model based on an adaptation of the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, specifically tailored for profiling the retail food environment of cities using administrative units as the spatial basis. This model leverages readily available open data sources, primarily OpenStreetMap (OSM) for identifying diverse establishment types and official administrative boundaries [7]. The core output of the model is a quantifiable profile, exemplified here by a modified RFEI (mRFEI), calculated for each administrative unit. The primary objective of this study is, therefore, to present this transparent, reproducible, and scalable CRISP-DM based model for assessing urban food environments using open geospatial data. To illustrate its practical application and evaluate its feasibility, the proposed model is applied in a case study of Madrid, the capital of Spain, where 131 administrative neighborhoods (known as "*barrios*" in Spanish) are profiled based on the mRFEI calculated from OSM data. The resulting spatial profile for Madrid demonstrates the model's utility in identifying geographic variations in food retail availability.

The remainder of this study is structured as follows: Section 2 provides an overview of the mRFEI concept and the use of OSM data. Section 3 details the proposed model, elaborating on each adapted CRISP-DM phase for this geospatial analysis context. Section 4 presents the main results of our study on Madrid. Finally, Section 5 presents conclusions regarding the model's applicability and discusses potential future work and refinement.

2. Preliminaries

The RFEI, created by the California Center for Public Health Advocacy, measures the ratio of fast-food restaurants and convenience stores (combined as the numerator) to supermarkets and produce vendors (the denominator). A modified version, known as the mRFEI, was later developed by the Centers for Disease Control and Prevention (CDC); the formula for its calculation is as follows [8]:

$$mRFEI = \frac{Number\ of\ healthy\ food\ retailers}{Number\ of\ all\ food\ retailers} x100\% \quad (1)$$

where healthy food retailers typically include supermarkets, larger grocery stores, warehouse clubs (e.g., Costco, Sam's Club), and produce stores within census tracts or half a mile from the tract boundary; and total food retailers include both healthy and less healthy options—such as supermarkets and produce vendors (healthy), as well as fast-food restaurants, convenience stores, and small grocery stores (often considered less healthy due to limited healthy options). The resulting score ranges from 0 to 100, with higher values indicating a healthier retail food environment.

Although the mRFEI has a standardized formula defined by the CDC to ensure consistency across geographic areas, researchers have adapted and expanded the way food access is measured to better capture real-world complexities. One common adaptation is adjusting the classification of food retailers considered "healthy" or "unhealthy," recognizing that store types and offerings vary by region and context [9]. Another important approach is the integration of socioeconomic and demographic data, which provides insight into how food environments interact with income levels, race, and transportation access [10]. These refinements offer a more nuanced and accurate representation of the food environment.

This study uses OSM data to identify different types of establishments and official administrative boundaries to analyze urban food environments. Considering the way OSM data is structured, the mRFEI formula has been adjusted. And this study focuses more on the ratio of healthy to unhealthy options; for example, if the ratio is 2 in each neighborhood, it means there are twice as many healthy options as unhealthy ones. More details are provided in Section 3.1.

3. Proposed model

This study utilizes a structured geospatial data analysis methodology to quantify and map the retail food environment within Madrid's administrative neighborhoods. OSM data provides the basis for identifying food outlets, which are analyzed within official neighborhood boundaries. A modified mRFEI is calculated as the primary metric. To ensure a systematic approach, the research methodology employed the CRISP-DM framework, adapted for geospatial tasks. This iterative framework guides the project through distinct phases from problem definition to deployment (Fig. 1).

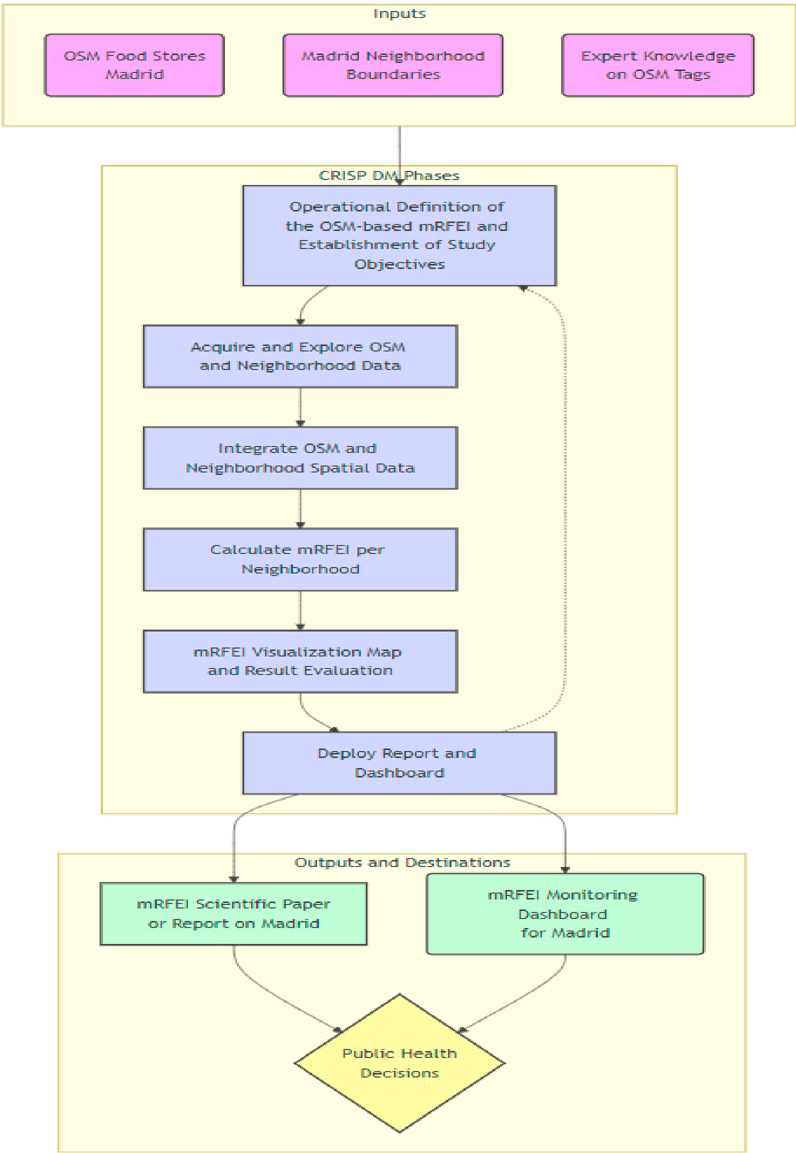

Fig. 1. Proposed Model

3.1. Operational Definition of the OSM-based mRFEI and Establishment of Study Objectives

This phase establishes both the conceptual framework and the concrete objectives of this study. Firstly, the key metric is defined: the Modified mRFEI. This definition is specifically constructed for the context of available data from OSM and guided by expert knowledge and the official documentation of the OSM Wiki [7]. Once this metric is defined, this phase establishes the research objectives focused on the city of Madrid (Spain).

The process to define the operational mRFEI involves the following fundamental tasks:

- OSM Tag Analysis: The *shop* and *amenity* tags documented in the OSM Wiki are systematically examined to identify those relevant to food retail outlets and related services in Madrid.
- Expert Classification Based on OSM Definitions: Applying expert knowledge of the local context and based on the standard definitions provided by the OSM Wiki, the selected OSM tags are classified into two categories ("Healthy", "Unhealthy"). The justification for each classification, linked to the official OSM definition, is detailed in Table 1.

Table 1. Classification of selected OSM tags into Healthy and Unhealthy Categories

OSM Tag	Description	Category	Classification Justification
shop=greengrocer	Shop primarily selling fresh fruit and vegetables.	Healthy	Specialized sale of fresh produce; clearly represents access to healthy options per OSM definition.
shop=farm	Shop selling produce (foodstuffs) directly from a farm.	Healthy	Implies sale of fresh primary products, often local, direct from the producer.
shop=supermarket	Large self-service store selling groceries and household goods.	Healthy	Although OSM doesn't define it by "healthy" and it sells mixed products, included as a primary source of access to diverse foods, including fresh items, for the general population.
amenity=marketplace	Area where goods, often food, are traded from stalls.	Healthy	Represents traditional or farmers' markets, significant sources of fresh and local produce.
shop=convenience	Small shop selling everyday items (snacks, drinks, basic groceries).	Unhealthy	Defined by OSM through the sale of impulse and convenience items, likely emphasizing less healthy or processed options.
shop=general	Shop selling a wide variety of diverse goods.	Unhealthy	OSM definition is broad, often rural. Included here assuming that in Madrid's urban context with specialized shops, 'general' stores primarily fill gaps with non-perishables, basics, and convenience items, functionally overlapping with `shop=convenience`.
shop=kiosk	Small stall selling items (newspapers, tobacco, sweets, snacks, drinks).	Unhealthy	Explicitly sells impulse and convenience products, generally considered less healthy, per the OSM definition.
amenity=vending_machine	Machine selling goods (often snacks, drinks).	Unhealthy	Automated source of convenience products, typically snacks and sugary drinks, well-defined in OSM.

- Formula Formalization: With the categories defined through the selected and justified OSM tags (Table 1), the mRFEI formula for this study is established:

$$mRFEI = \frac{Number\ of\ OSM\ establishments\ with\ \text{“Healthy”}\ tags}{Number\ of\ OSM\ establishments\ with\ \text{“Unhealthy”}\ tags} x100\% \quad (2)$$

Once the OSM-based mRFEI indicator is established, the general objective is defined as using this defined mRFEI to characterize and spatially profile the retail food environment of Madrid's administrative neighborhoods, identifying and visualizing the geographic heterogeneity in the relative availability of food options. Furthermore, the following specific objectives are established:

- Acquire and validate the official geographic boundaries of Madrid's neighborhoods, establishing the spatial unit of analysis.
- Implement the calculation of the defined mRFEI for each of the neighborhoods.
- Generate tools with interactive cartographic visualizations that show the spatial distribution of the mRFEI and identify the food patterns of each neighborhood.

- Evaluate the feasibility of the approach and the utility of the calculated index as a tool for analyzing the urban food environment.

3.2. Acquire and Explore OSM and Neighborhood Data

This second phase consisted of the initial acquisition and exploration of the key geospatial data.

Regarding acquisition, the official shapefile (SHP) with the polygon geometries and attributes of Madrid's neighborhoods was obtained [11]. Furthermore, elements (points) corresponding to the shop and amenity tags previously defined as "Healthy" and "Unhealthy" for the Madrid area were extracted from OSM [7].

For exploration, firstly, the structure of both datasets was examined, validating both the spatial and tabular information. Regarding spatial information, a fundamental discrepancy was found in the Coordinate Reference Systems (CRS): the neighborhood data used a local projected system (UTM zone 30N) with mercator, while the OSM data was in the global geographic system WGS84 which uses longitude and latitude. This difference therefore requires subsequent harmonization. The tabular (non-spatial) data of both datasets were evaluated, determining them to be of sufficient quality to continue this study.

3.3. Integrate OSM and Neighborhood Spatial Data

Continuing the methodological process, this third phase focused on the exhaustive preparation and spatial integration of the acquired datasets, in order to generate a unified database ready for the calculation of the mRFEI at the neighborhood level. This process required several fundamental transformations and geospatial operations to ensure the quality and compatibility of the information.

Initially, the dataset extracted from OSM was processed to exclusively select the point geometries representing individual establishment nodes, thus discarding other geometry types. Subsequently, applying the classification criteria established in Phase 1 (detailed in Table 1), each of these selected OSM points was programmatically categorized as "Healthy" or "Unhealthy", adding this essential attribute information to each record.

A crucial technical step was the harmonization of CRS. Given the previously identified discrepancy between the global geographic system (WGS84) of the OSM data and the local projected system (UTM zone 30N) of the Madrid neighborhood boundaries, a reprojection of the neighborhood polygons to the WGS84 system was performed, thereby ensuring that both datasets shared the same standard global CRS. This spatial alignment was indispensable for the correct execution of subsequent geospatial operations.

Finally, spatial integration was carried out through a spatial join operation. This procedure geographically linked each classified establishment point (Healthy/Unhealthy, already in WGS84) with the corresponding administrative neighborhood polygon (now also in WGS84) that spatially contained it. As a result of this join, the unique neighborhood identifier (*GEOCODIGO*) was attributed to each point record. Upon concluding this phase, a geospatial dataset was available where each relevant establishment was classified and associated with the neighborhood to which it belongs (all in WGS84), ready for the index calculation phase.

3.4. Calculate mRFEI per Neighborhood

This fourth phase implemented the core quantitative analysis: the calculation of the mRFEI for each defined spatial unit, the administrative neighborhoods of Madrid. The process relied on the spatially integrated and prepared data from the preceding phase.

First, an aggregation of the establishment data at the neighborhood level was performed. Using the unique identifier (*GEOCODIGO*) assigned to each retail outlet point during the spatial join phase, all records corresponding to the same neighborhood were grouped. Within each group (neighborhood), the fundamental metrics required for the index were calculated: the total number of establishments classified as "Healthy" (*healthy_count*) and the total number of establishments classified as "Unhealthy" (*unhealthy_count*), achieved by summing the previously assigned categorical indicators.

With the counts of healthy and unhealthy establishments available for each neighborhood, the calculation of the mRFEI index proceeded. The formula *mRFEI* = (*healthy_count* / *unhealthy_count*) * 100 was applied to each neighborhood record.

Finally, the results of this aggregation and calculation (the *healthy_count*, *unhealthy_count*, and the *mRFEI* index itself) were integrated, via a merge operation based on the *GEOCODIGO*, into the original geographic dataset of neighborhood polygons. This produced the final dataset for analysis and visualization, where each neighborhood polygon now contained, in addition to its original information, the calculated metrics concerning its retail food environment (see Table 2.).

Table 2. Extract of Calculated mRFEI and Outlet Counts per Madrid Neighborhood (spatial data are not shown)

GEOCODIGO	neighborhood	healthy_count	unhealthy_count	mRFEI
079011	011 Palacio	11	86	12.79
079012	012 Embajadores	33	55	60
079013	013 Cortes	9	16	56.25
079014	014 Justicia	13	25	52
079015	015 Universidad	27	115	23.48

3.5. mRFEI Visualization Map and Result Evaluation

This phase centered on the visual representation of the mRFEI calculation results and the critical evaluation of both the generated map and the index values themselves, assessing their meaning and adequacy concerning the study's objectives.

The primary activity was the generation and analysis of interactive cartographic visualization. This map presented the spatial distribution of the mRFEI using thematic mapping of the neighborhoods, employing the previously defined dual color scale (<100 i.e. signifying a relative abundance of "Unhealthy" outlets, and ⩾100 i.e. relative abundance of "Healthy" stores) to facilitate interpretation. Simultaneously, point layers corresponding to establishments classified as "Healthy" and "Unhealthy" were overlaid, allowing for detailed visual inspection of the underlying food environment composition within each neighborhood. Emerging spatial patterns on the map were carefully examined, seeking to identify geographic clusters, gradients, or anomalies in the index distribution. This interactive map is shown in Fig. 2.

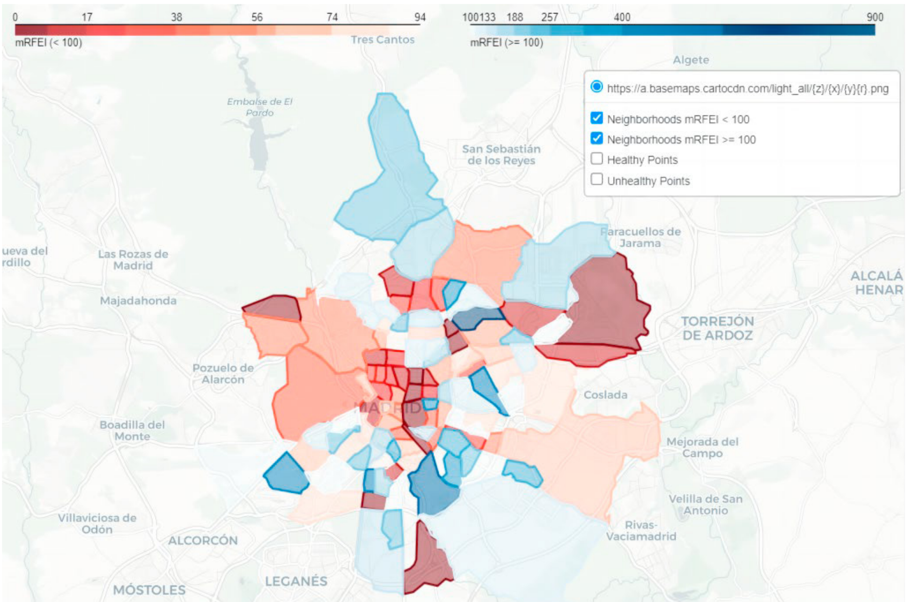

Fig. 2. Spatial Distribution of the mRFEI in Madrid Neighborhoods.

Parallel to the visual map evaluation, the numerical mRFEI results were analyzed. This included studying the statistical distribution of the index (range, mean, median) and identifying neighborhoods with extreme or undefined values. The consistency between the map's visual representation and the underlying numerical values was assessed to ensure the validity of the interpretation.

Finally, a comprehensive evaluation of the process and results was conducted. The implications of limitations identified in previous phases (OSM data quality, index simplifications) on the reliability and interpretability of the findings and the mRFEI map were considered. This evaluation determined the suitability of the results for meeting the initial objectives and guided the decision regarding the need for potential iterations or refinements before the final deployment phase.

3.6. Deploy Report and Dashboard

This concluding phase involved preparing and sharing the project's results. Key activities included finalizing this scientific report, which details the methodology and findings, and potentially developing an interactive monitoring dashboard to visualize the calculated mRFEI for Madrid. These materials serve to communicate the assessment and support subsequent public health or planning decisions.

4. Results

In our case, the spatial analysis of the mRFEI across Madrid reveals a distinct geographic pattern concerning the retail food environment. A general tendency observed is that the central and often more touristic neighborhoods tend to exhibit lower mRFEI scores (predominantly values well below 100, indicated by red shades on the map shown in Fig. 2), suggesting a relative predominance of "Unhealthy"-classified outlets compared to "Healthy" ones. Examination of the underlying counts (provides examples) often shows high numbers of convenience stores, kiosks, or general stores in these central "barrios" (e.g., Palacio, Universidad, Recoletos, Almagro), which drives the denominator of the index up relative to the count of supermarkets or greengrocers. Conversely, several peripheral and primarily residential neighborhoods display significantly higher mRFEI scores (values $\geqslant 100$, indicated by blue shades), often resulting not necessarily from a high absolute number of "Healthy" outlets, but rather from a markedly lower count of "Unhealthy"-classified establishments (e.g., Zofío, Entrevías, Canillejas show few unhealthy outlets relative to healthy ones). While notable exceptions exist, with some peripheral areas also showing low scores, the overall spatial distribution highlighted by the mRFEI suggests significant heterogeneity, with central, high-traffic areas generally profiling as less "healthy" according to this specific retail-based metric compared to several, though not all, residential areas further from the city center.

These findings are helpful for urban planning and public health policy. Policymakers could consider targeted zoning regulations or incentive programs to encourage the establishment of healthier food outlets in these central areas. For example, subsidies or relaxed permitting for greengrocers or fresh food markets could help rebalance the food environment. Conversely, in residential peripheries where unhealthy food options are less prevalent, maintaining or enhancing the accessibility of healthy outlets should be prioritized to preserve and support these relatively favorable conditions. Furthermore, integrating the mRFEI into broader urban health strategies could enhance local governance decisions regarding food deserts, community health disparities, and even transportation planning. Overall, the spatial heterogeneity captured by the mRFEI is not only a diagnostic tool but also a foundation upon which more health-oriented urban food policies can be built.

5. Conclusions and future work

This study successfully demonstrated a replicable methodology using OSM data to calculate and map the mRFEI across Madrid's neighborhood. It proved effective in identifying significant spatial heterogeneity in the distribution of food retail types, offering a useful view of urban food environments. The results indicate a general trend where central areas exhibit relatively lower mRFEI scores compared to several peripheral residential areas, suggesting potential spatial disparities in the retail food environment that may be shaped by underlying socioeconomic and urban structural factors. While highlighting the utility of open data for such assessments, limitations concerning

OSM data quality and index simplification must be acknowledged. Future work should focus on validating these findings, potentially refining the mRFEI calculation by incorporating additional outlet types or weighting schemes and integrating socio-economic variables to explore equity dimensions and better inform targeted public health and urban planning strategies.

Acknowledgements

The authors would like to acknowledge the financial support from the Spanish Ministry of Science and Innovation (grant number PID2022-139297OB-I00). They would also like to thank the team at the Instituto de Investigación Científica (IDIC) of the Universidad de Lima for their support and guidance.

References

[1] Pineda, E., Brunner, E. J., Llewellyn, C. H., and Mindell, J. S. (2021) "The retail food environment and its association with body mass index in Mexico." *International Journal of Obesity* **45 (6)**: 1215–1228.

[2] Pineda, E., Stockton, J., Scholes, S., Lassale, C., and Mindell, J. S. (2024) "Food environment and obesity: a systematic review and meta-analysis." *BMJ nutrition, prevention & health* **7(1)**: 204-211.

[3] Bernsdorf, K. A., Bøggild, H., Aadahl, M., and Toft, U. (2024) "Measuring associations between the food environment and dietary habits: Comparing the proportion and density of food outlets." *BMC Public Health* **24 (1)**: 3445.

[4] Guibrunet, L., Hoekman, P., Bortolotti, A., and Battersby, J. (2024) "Data requirements for a systematic analysis of urban food flows and their sustainability outcomes." *Environment and Planning B*, 23998083241298431.

[5] Glover, B., Mao, L., Hu, Y., and Zhang, J. (2022) "Enhancing the Retail Food Environment Index (RFEI) with Neighborhood Commuting Patterns: A Hybrid Human−Environment Measure." *International Journal of Environmental Research and Public Health* **19 (17)**: 10798.

[6] Pineda, E., Stockton, J., and Mindell, J. S. (2024) "The Retail Food Environment Index and its association with dietary patterns, body mass index, and socioeconomic position: A multilevel assessment in Mexico." *PLOS Global Public Health* **4(10)**, e0003819.

[7] OpenStreetMap Wiki Contributors. Map Features. Retrieved from https://wiki.openstreetmap.org/wiki

[8] CDC. Census tract level state maps of the modified food environment index (mRFEI). Retrieved from https://stacks.cdc.gov/view/cdc/61367

[9] Vandevijvere, S., Mackenzie, T., and Mhurchu, C. N. (2017) "Indicators of the relative availability of healthy versus unhealthy foods in supermarkets: A validation study." *International Journal of Behavioral Nutrition and Physical Activity* **14 (1)**: 53.

[10] Kemp, C., Collins, J., and Palermo, C. (2019) "Is the type and location of grocery stores a predictor of healthy and unhealthy food availability? A cross-sectional study." *Nutrition & Dietetics: The Journal of the Dietitians Association of Australia* **76 (3)**: 277–283.

[11] Comunidad de Madrid. Nomenclátor Geográfico de Municipios y Entidades de Población. Retrieved from https://gestiona.comunidad.madrid/nomecalles_web/