BUSINESS OPPORTUNITIES ANALYSIS USING GIS: THE RETAIL DISTRIBUTION SECTOR

Norat Roig-Tierno

Department of Economics and Social Sciences, Universitat Politècnica de València, Valencia, Spain

Amparo Baviera-Puig

Department of Economics and Social Sciences, Universitat Politècnica de València, Valencia, Spain

Juan Buitrago-Vera

Department of Economics and Social Sciences, Universitat Politècnica de València, Valencia, Spain

Abstract:

The retail distribution sector is facing a difficult time as the current landscape is characterized by ever-increasing competition. In these conditions, the search for an appropriate location strategy has the potential to become a differentiating and competitive factor. Although, in theory, an increasing level of importance is placed on geography because of its key role in understanding the success of a business, this is not the case in practice. For this reason, the process outlined in this paper has been specifically developed to detect new business locations. The methodology consists of a range of analyses with Geographical Information Systems (GIS) from a marketing point of view. This new approach is called geomarketing. First, geodemand and geocompetition are located on two separate digital maps using spatial and non-spatial databases. Second, a third map is obtained by matching this information with the demand not dealt with properly by the current commercial offer. Third, the Kernel density allows users to visualize results, thus facilitating decision-making by managers, regardless of their professional background. The advantage of this methodology is the capacity of GIS to handle large amounts of information, both spatial and non-spatial. A practical application is performed in Murcia (Spain) with 100 supermarkets and data at a city block level, which is the highest possible level of detail. This detection process can be used in any commercial distribution company, so it can be generalized and considered a global solution for retailers.

Keywords:

Business opportunities, retailers, location, strategy, GIS, geomarketing.
**INTRODUCTION**

It has always been said that the key variables of a successful retail distribution company are location, location, location, so from this statement it is easy deduce the importance of a proper location strategy for retailers (Ghosh and McLafferty 1982). In retail companies, the opening of a new store or outlet carries an inherent risk because of the high monetary costs associated. Also, a store that is unsuccessful due to a poor choice of location can have a significant negative impact on the image of the company. Consequently, the analysis of location is vital for retail and commercial enterprises (Hernández and Bennison 2000).

If the process of choosing the site of a retail establishment has always been complicated, this is now truer than ever, since environmental circumstances have recently worsened. The current situation is characterized by ever-increasing competition, resulting in lower margins and the exploitation of all possible market segments. Any element of competition that can triumph in this environment has a very high value (Clarke 1998). The search for an optimal location strategy has the potential to become the differentiating factor. This is the reason why developing methodologies and processes that identify new business locations is essential.

Geography is being handed an increasing importance for its key role in the current understanding of the success of a business (Alcaide et al. 2012). Nevertheless, Wood and Reynolds (2012) point out that although the literature places great importance on spatial analysis and geographic information in decision-making in the location strategy of companies, this is not the case in practice. Thus, there are possibilities to develop methodologies for determining new commercial locations by integrating such spatial and geographic information with clear, practical applications.

Therefore, the objectives of this research are as follows.

1. First, to use statistical information together with spatial components to identify demand and competition. Given the combination of the statistical and spatial components in this identification process, these concepts are referred to as geodemand and geocompetition.

2. Second, to determine new locations for commercial establishments by jointly analyzing geodemand and geocompetition.

To do this, the commercial distribution sector of frequently purchased products is analyzed in a given Spanish city (Murcia).

The paper is structured as follows. First, the theoretical framework on which this research is based is presented. Second, the methodology used in the analysis is developed. Then, the results of the study are shown and discussed, and, finally, the main conclusions from the results are drawn.

**THEORETICAL FRAMEWORK**

Numerous authors have contributed in different ways to the methodology used by retailers to determine the location of new establishments. Church (2002) states that the success of many future applications of the location of retail outlet sites may be closely linked to Geographical Information Systems (GIS) due to the fact that these systems are responsible for working with spatial information. For this reason, the use of GIS as a site evaluation tool is presented, as well as
different definitions derived from its implementation (geomarketing, geodemands, geocompetition and trade area).

**Site evaluation tools: the use of GIS**

In complex decision processes involving a wide variety and volume of information, along with a substantial subjective component, visualization methods are extremely useful. Therefore GIS, based on matching digital maps with relational databases, will undoubtedly be essential for the future development of the decision processes for locating establishments (Mendes and Themido 2004).

While it is true that location theory and GIS have evolved almost independently, they currently support each other together with research into operations. This research studies models for decision-making in which the techniques are equally applicable in areas of spatial and non-spatial character (Church and Murray 2009).

In fact, the use of GIS by companies and organizations is increasing (Rob 2003; Chen 2007). In recent years, various associations between decision-making models and GIS packages have been made. Harris and Batty (1993) and Birkin et al. (2002) explore the different possibilities that these technologies can offer to solve the problems of planning and locating retail outlets. Such associations between applications are known as loosely coupled, whereas solutions that include the functionality of decision-making programming within the actual GIS packages are known as strongly coupled. This strategy is based on the acceptance that there is no single software tool or technology that can meet the needs of planners and, therefore, they will have to adapt the current (and future) technologies to satisfy their needs (Harris and Batty 1993).

Instead, Murad (2003) defines the trade areas of two shopping centers and calculates its share of market penetration using various tools exclusively offered by GIS. Later, Murad (2007) evaluates the sites of shopping centers in relation to the customers’ location, again using only GIS. The research of Clarkson et al. (1996) and Hernández and Bennison (2000) in the United Kingdom confirms the integration of GIS within the usual processes of location decision of firms, without introducing new methodologies, but rather by changing the approach that has been used up to now. The visualization of the data and results in such a complex decision process is one of the reasons for its success (Hernandez 2007; Ozimec et al. 2010). Thus, the use of GIS has facilitated the understanding of geographical information for managers who lack expertise, helping them to make important yet difficult decisions. Furthermore, the latest advances in GIS allow technicians to define, monitor and automate the calculations and the creation of maps necessary for the resolution of a problem using flow charts (Suárez-Vega et al. 2012). These charts greatly improve the ease of work because the whole process is better understood and it reduces the working time when processing repetitive tasks. This is often the case when evaluating different site locations over and over again.

**GIS and geomarketing**

With the integration of GIS in decision-making processes, the spatial variable takes a relevant role as a descriptive and explanatory variable. In fact, Sleight et al. (2005) state that people who share
geographic environments also tend to share behaviors, consumption habits and related attitudes. The location of customers and the analysis of their environment become especially relevant. The integration of GIS in the study and analysis of the customer, both from a spatial and non-spatial perspective, opens the way for a new field of study called geomarketing (Baviera-Puig et al. 2009). Latour y Le Floc’h (2001) define the term geomarketing as an integrated system for data processing software, and statistical and graphical methods designed to produce useful information for decision-making, through instruments that combine digital maps, graphs and tables. Meanwhile, Chasco (2003) states that geomarketing is a set of techniques for analyzing the economic and social reality from a geographical point of view, through mapping tools and spatial statistics tools.

From a more sociological viewpoint, Alcaide et al. (2012) argue that geomarketing is the area of marketing that is aimed at global customer knowledge, and the needs and behaviors of customers within a given geographical area. All this information helps the company to have a more complete view of its customers and to identify their needs. Finally, Baviera-Puig et al. (2013) suggest that geomarketing can be defined as the discipline that uses GIS as a tool for analysis and decision-making in marketing, in order to meet the needs and wants of consumers in a profitable way for the company.

Geomarketing involves the following elements: databases, cartographic information and GIS for processing and managing information. Databases can be either internal (sales, corporate data, customers, etc.) or external (statistical institutes, municipal census, Chambers of Commerce, etc.). The digital maps can come from several sources, such as private companies (TeleAtlas, Navteq, etc.) or Cartographic Institutes. For example, in Spain the National Institute of Statistics (Instituto Nacional de Estadística, INE), Electronic Office of Cadastre (Sede Electrónica del Catastro) and the National Centre for Geographic Information (Centro Nacional de Información Geográfica, CNIG) all manage and provide digital maps. Finally, GIS is the tool in charge of linking databases with geographic information. As a result, data processing can be performed with this tool and any studies or analysis required may be carried out.

There are two new concepts that can be derived from those defined above: geodemand and geocompetition. The former, geodemand, can be defined as the location on a digital map of the customers of a product or a service in a particular market. On the other hand, geocompetition is the location on a digital map of the competitors of a business, and the delineation of their trade area in a particular market.

Another important concept within spatial analysis is the trade area or market area of a retail establishment. It can be defined as the geographic area in which the retailer generates all of its sales during a specific period (Applebaum and Cohen 1961). Baviera-Puig et al. (2012) define the trade area of a retailer as the geographic area in which the company is able to attract customers and generate sales.
METHODOLOGY

This research uses three different techniques but they all have the support of GIS. First, geodemand analysis is performed to locate the customers of a product or a service on a digital map. Second, competition is also located on a digital map as it is one of the main factors to consider when determining the site of a new store. This process is known as geocompetition analysis. The third and final technique is to define the possible sites for new stores considering all the previous steps as a whole.

Data

This study is conducted in the city of Murcia (Spain) and analyzes the commercial distribution sector of frequently purchased products.

Using the INE (2011) as a source of information, Murcia has a population of 442,203 inhabitants divided into 386 census tracts. A census tract is a territorial unit which is defined according to operational criteria for fieldwork in statistical operations depending on its population size. In Murcia, the average population of a census tract is 1,146 inhabitants, with a minimum of 650 and a maximum of 2,630 inhabitants.

From the Electronic Office of Cadastre (2012), in Murcia there are 194,615 housing units in 48,748 city blocks. The sum total of the surface area of these city blocks is 20,888,346.37 m², the average size being 428.49 m². Only those city blocks with residential use are considered.

According to Nielsen (2012), Murcia has 100 supermarkets from 23 different chains (Table 1).

Table 1. Supermarkets in Murcia

<table>
<thead>
<tr>
<th>CHAIN</th>
<th>N° OF SUPERMARKETS</th>
<th>% SUPERMARKETS</th>
<th>SUM TOTAL OF NIELSEN SURFACE (m²)</th>
<th>% SURFACE</th>
<th>AVERAGE SURFACE (m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>3.00</td>
<td>2600</td>
<td>1.95</td>
<td>866.67</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3.00</td>
<td>1450</td>
<td>1.08</td>
<td>486.33</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3.00</td>
<td>2776</td>
<td>2.08</td>
<td>925.33</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1.00</td>
<td>375</td>
<td>0.28</td>
<td>375.00</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>10.00</td>
<td>4613</td>
<td>3.45</td>
<td>461.30</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1.00</td>
<td>800</td>
<td>0.60</td>
<td>800.00</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>4.00</td>
<td>8270</td>
<td>6.19</td>
<td>2067.50</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1.00</td>
<td>9108</td>
<td>6.82</td>
<td>9108.00</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>2.00</td>
<td>22531</td>
<td>16.86</td>
<td>11265.50</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>4.00</td>
<td>1963</td>
<td>1.47</td>
<td>490.75</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>2.00</td>
<td>20677</td>
<td>15.47</td>
<td>10338.50</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>6.00</td>
<td>3396</td>
<td>2.54</td>
<td>566.00</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1.00</td>
<td>750</td>
<td>0.56</td>
<td>750.00</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>2.00</td>
<td>699</td>
<td>0.52</td>
<td>349.50</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1.00</td>
<td>1050</td>
<td>0.79</td>
<td>1050.00</td>
</tr>
</tbody>
</table>

**Geodemand analysis**

The steps for geodemand analysis are twofold: (i) calculate the number of housing units per city block from cadastral data; and (ii) estimate the average number of people per city block.

To calculate the number of housing units per city block, alphanumeric data of the city blocks from the cadastral database of the municipality is associated to the graphical data with the help of GIS. Next, this data is matched with the number of people from the Municipal Census (Padrón Municipal).

As a result, an estimate of people living in each city block is obtained. This step is more complicated because the information from the Municipal Census refers to the census tract and, therefore, contains several city blocks. The process for calculating the number of people per household in each census tract is as follows.

(i) First, obtain the Municipal Census data from the INE, along with the census tracts, for the municipality in question so that this information can subsequently be integrated in the GIS.

(ii) Second, the residents within each census tract are grouped into the housing units present in that tract.

**Geocompetition analysis**

The steps to perform geocompetition analysis are threefold: (i) identify the competition; (ii) locate competitors on a digital map; and (iii) define their trade areas.

First, all the supermarkets which are in the city are selected in order to be geolocated. The geolocation consists of providing coordinates (x,y) for the addresses of the establishments. This task is performed entirely by GIS.

Once the competitors have been geolocated, the trade areas of each supermarket are calculated and delineated. The trade area is defined in terms of the surface area or size of the establishment. The surface area or size is a determining factor in calculating the trade area of a supermarket (Reilly 1931; Huff 1963). According to Table 2, a supermarket with a surface area of 800 m² corresponds to an isochrone of 8 minutes, equivalent to a maximum distance traveled of 533 meters. This
maximum distance increases if the surface area or size of the supermarket is greater, and it decreases if the surface area is smaller.

Table 2. Trade areas of supermarkets according to their size

<table>
<thead>
<tr>
<th>SURFACE AREA OR SIZE OF SUPERMARKET (m²)</th>
<th>MAXIMUM DISTANCE TRAVELED (m)</th>
<th>ISOCHRONE (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>300&lt;S ≤ 600</td>
<td>333</td>
<td>5'</td>
</tr>
<tr>
<td>600&lt;S ≤ 1000</td>
<td>533</td>
<td>8'</td>
</tr>
<tr>
<td>1000&lt;S ≤ 2500</td>
<td>667</td>
<td>10'</td>
</tr>
<tr>
<td>S&gt;2500</td>
<td>1200</td>
<td>18'</td>
</tr>
</tbody>
</table>

When two or more trade areas overlap due to the proximity between two or more supermarkets, the resulting area acquires the power of attraction of the sum of the corresponding trade areas. Potential customers who live in that area have a higher commercial offer; therefore, this area is more saturated or occupied. In the literature, this process is known as cannibalization because supermarkets fight to get those customers (Kelly et al. 1993). In other words, cannibalization is understood as the portion of the market share of the new store captured from the existing stores (Suárez-Vega et al. 2012).

Kernel density estimation

Kernel density is a non-parametric estimation of the density function of a random variable (Rosenblatt 1956). The basic concept of spatial density refers to a relationship between the level of presence of such a variable in a given zone and the surface area of the zone. The final result is expressed in units of the particular phenomenon per unit area. Conceptually, the Kernel density function aims to calculate the density of points in a given area according to the distance between the points, as long as the points have the same weight. However, different weights can be assigned to each point in order to attribute different points with greater or lesser relative importance.

Although there exist different types of models when using Kernel estimators (Moreno 1991), Härdle (1991) claims that the choice of model is almost irrelevant for the quality of the estimate and, consequently, for the final outcome of the analysis. Therefore, in this study the quadratic Kernel function described by Silverman (1986) is used. It is also the estimator integrated in the GIS employed to conduct the research.

Generically, and for the univariate case, the estimator can be written as follows:

\[
\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x - x_i}{h} \right)
\]

Where:

\(\hat{f}(x)\) = estimator of the Kernel density function.

\(x\) = point where density is estimated.

\(x_i\) = value of the variable in the case \(i = 1, \ldots, n\).
\( h \) = window width or smoothing parameter. This parameter limits the influence of each datum to a field, defined by the window. As the width of the window is bigger, it causes an increased smoothing in the resulting map. The choice depends on the purpose of the study.

\( K \) = Kernel symbol. In the case of the Kernel quadratic function:

\[
K = \frac{3}{\pi} (1 - u^2)^2, \text{ for } |u| \leq 1, \text{ where } u = (x - h_i)/h \]

More specifically, in this research, based on Moreno (2007), the unit of study is the pixel. The pixel is a square of size designed to represent a portion of the space on the digital map, and which is associated with a value or color, in turn associated with the element represented in that portion of territory. In this new scenario, a circular environment is defined for every pixel in the map and is then used as a baseline. The centroid of each pixel is the centre of the circle and the points that fall within it are used to form the dividend. Each point is weighted unevenly, according to its proximity to the centroid of the pixel: the pixels nearer the centroid have a greater weight than those farther away. The expression used is:

\[
L_j = \sum_{i \in C_j \cap \mathbb{R}^{2}} \frac{3}{\pi} \left(1 - \frac{d_{ij}^2}{r^2}\right)^2
\]

Where:

- \( L_j \) = density estimated in the pixel.
- \( d_{ij} \) = distance between points \( i \) and \( j \).
- \( r \) = width of the window or search radius, which determines the degree of smoothing.
- \( C_j \) = \( \{i | d_{ij} < r\} \), so that the set is formed by the \( i \) points whose distance to the centroid of the pixel \( j \) is smaller than the radius of the circle prescribed.

In this research, a pixel size of 5 meters has been defined. Such a small value is chosen because the smallest unit of reference is the city block and, therefore, a larger pixel size would cover several city blocks. A search radius of 300 meters has also been established, considering the average radius of a trade area.

**RESULTS AND DISCUSSION**

**Geodemand and geocompetition identification**

One of the big questions that geomarketing aims to answer is where the potential business clients are. Based on the analysis of geodemand, and once the number of persons per household in each census tract has been estimated, a map of the population density of the study area is generated. Figure 1 represents the commercial demand of a portion of the city of Murcia on a 3D map. It appears that the high peaks coincide with an accumulation of population, so the greater the number of people that there are at one point, the higher the peak.

**Fig. 1 Geodemand of a portion of the city of Murcia**
Nevertheless, competition is one of the main factors to consider when determining the site of a new store. For this reason, it is important to conduct a detailed analysis to determine geocompetition in the city of Murcia in a quantitative and visual way. Areas with a higher retail space, previously called saturated or occupied, can then be detected. These areas are those that are more or less covered by a trade area of some establishment. Similarly, those areas with a lower retail space or with low commercial offer are also determined. As shown in Figure 2, there is a high saturation in the centre of the city as this is where most trade areas of different stores overlap.

**Fig. 2 Geocompetition of a portion of the city of Murcia**

-Determination of business opportunities-

Davis (2006) describes how the spatial dispersion of consumers and sellers influences market shares and substitution patterns between different retail options. Therefore, this spatial dispersion can also contribute to determining site locations for new stores. The term business opportunity refers to an area that offers a high population density and little or no commercial competition.
Consequently, business opportunities translate as potential site locations for new commercial establishments.

To identify such opportunities in the city of Murcia, geodemand and geocompetition are analyzed jointly.

(i) After matching information resulting from both analyses, a third map is obtained that shows, on the one hand, the population free from commercial offer and, on the other hand, areas with a low commercial offer.

(ii) Following this, Kernel density analysis is carried out for this third map to determine which areas have a greater concentration of potential customers and how many potential clients can be found. In Figure 3, the darker areas correspond to a higher population density with little access to commercial offer; i.e. not adequately served by the potential commercial offer available.

Fig. 3 Business opportunities analysis

After the Kernel density estimation, a study and a description of the features of the areas most likely to host a new retail site are performed. In this case, areas A, B and C are investigated in depth, as they are the points that form the output of the study. The location of these three areas appears in Table 3. As a result, one or two of them can be selected or they can be prioritized in order to open a new store.

Table 3. Location of the areas that form the output of the study

<table>
<thead>
<tr>
<th>AREA</th>
<th>COORDINATES OF THE CENTER* (m)</th>
<th>AXES OF THE AREA (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>A</td>
<td>664415.36 4205893.97</td>
<td>406.33 195.85</td>
</tr>
<tr>
<td>B</td>
<td>664920.28 4204419.34</td>
<td>407.82 300.56</td>
</tr>
<tr>
<td>C</td>
<td>666780.34 4206937.11</td>
<td>419.36 295.78</td>
</tr>
</tbody>
</table>

*Projected Coordinate System ETRS 1989 UTM Zone 30 N.
Managerial implications

With the proposed methodology, the aim is to determine the most appropriate site when opening a new store in order to conduct a proper location strategy. The benefits of an appropriate location strategy are manifold (Ghosh and McLafferty 1982). These benefits include a better distribution of the financial resources of the company (Alarcón 2011), increased competition (Clarke 1998), a better image for consumers and, of course, the survival of the company. This methodology can be used in any commercial distribution company, regardless of the industry, and can therefore be considered a global solution for retailers. While Murad (2007) uses GIS to assess the sites of shopping centers in relation to the location of customers, the study only considers two malls. However, the current research takes into account 100 different supermarkets and population data at the city block level, representing the highest degree of detail. This is possible because GIS is able to handle large amounts of information. In addition, GIS rules out sites depending on the land use or because potential commercial sites do not meet the minimum size requirements. Despite handling so much information and having to perform a large number of calculations, all these processes can be automated and visualized using flowcharts offered by GIS. With Kernel density, a representation of the trend or overall pattern of distribution of the population with no commercial offer, or with very little commercial offer, is obtained. The final product is an isopleth, or isodensity map, for easy viewing of the results and to facilitate decision-making by managers, regardless of their professional background (Musyoka et al. 2007).

CONCLUSIONS

The objective of this research consists of creating a process that contributes to the definition of the location strategy for retailers. To do so, GIS tools were used as a means of exploiting alphanumeric and spatial databases from the viewpoint of geomarketing. Under the first objective, the geodemand and the geocompetition are identified on a digital map and geodemand is then calculated, quantified, and visualized for the city of Murcia. As a result, by using GIS, it has been possible to determine the distribution of potential customers in this area. It can be observed here that the population is not evenly distributed in the city, as there is a series of peaks that show high concentrations of people; i.e. potential markets. The geocompetition analysis quantitatively and visually determines the areas of the city of Murcia with the highest retail space (saturated or occupied areas) and those areas not served by the fully available current commercial offer. Regarding the second objective, by creating an intersection between the geodemand and the geocompetition, those areas of Murcia where geodemand is unmet by the geocompetition can be detected. Thanks to Kernel density estimation, these areas can be distinguished in a quantitative and visual way. These locations are the main output of the study and correspond to possible sites for new commercial establishments.

As future research, the analysis performed with GIS can be complemented with multicriteria decision techniques. These techniques can be discrete if the decision alternatives are finite, or multiobjective when the problem takes an infinite number of possible alternatives (Thaler 1986).
In the case of this particular study, the former would be chosen as the results uncovered here are of a finite number of options. According to Berumen and Llamazares (2007), the main discrete multicriteria decision methods are linear weighting (scoring), Multiattribute Utility (MAUT), overrating relations and Analytic Hierarchy Process (AHP).

ACKNOWLEDGMENTS

The authors would like to thank the *Consum-Universitat Politècnica de València* Chair (Cátedra) for its collaboration in this study.

REFERENCES


