A GIS Solution to Prioritizing Site Selection Efforts in Chicago, Illinois

Allison L. Kurth
Department of Resource Analysis, Saint Mary's University of Minnesota, Minneapolis, MN 55404

Keywords: GIS, Sales, Grids, Opportunity, Potential, Site Selection

Abstract

In past decades, the precedent for businesses looking for new retail sites has been to choose locations based on intuition or personal experience. However, in today's struggling economy and crowded market place, many businesses are looking for more analytical methods. This research provides a geographic method for prioritizing potential new retail locations. By using demographic data and Geographic Information Science (GIS), the time salespeople spend pursuing unsuitable store location will be reduced. This research focuses on the Chicago, Illinois seven county metro area, but these same methods could be applied across the United States, perhaps across the world.

Introduction

Significance of Research

In the past, retailers have relied on emotionally subjective methods (Simkin, Doyle, and Saunders, 1985) or intuition (Rogers, 1987) for choosing new retail locations. Two of the main reasons for the lack of analytical analysis in this area are high expense and not enough consumer data. A paradigm switch has occurred in the strategy retailers use to increase their number of outlets. In the past, a retailer would consider only where to open a new store. In the last decade, the approach has become more holistic. Retailers consider not only new locations, but the bigger picture by looking at expanding or moving current locations. Companies look at the sales territory of each location and consider ways to expand to cover as much of the targeted population as possible (Hernández and Bennison, 2000).

It is commonly accepted across all retail industries that location is a key factor crucial to success. As noted in Kimes and Fitzsimmons (1990), a business that can choose new sites quickly and correctly has a significant competitive advantage. This research will detail a method for combining demographic data, proprietary business knowledge and consumer tracking resources into an analytical method for prioritizing which locations in Chicago the sales team should consider top priority. This will save salespeople time and therefore money. There are over ten thousand potential retail locations for this company; if GIS can reduce that list to a more manageable number of prospects that have high potential, the salespeople should become more efficient and effective.

These data and this location are ideal because Chicago is a populous and vast city that should showcase the strengths of this type of analysis. This study will not disclose the company or industry due to the proprietary nature of sales data. However, this company's method for determining the Remaining Opportunity (RO) is similar to industry standards and should represent a good

Kurth, Allison. 2012. A GIS Solution to Prioritizing Site Selection Efforts in Chicago, Illinois. Volume 14, Papers in Resource Analysis. 10 pp. Saint Mary's University of Minnesota Central Services Press. Winona, MN. Retrieved (date) from http://www.gis.smumn.edu.

example case. Indeed, this methodology could be applied to other companies with retail location decisions to make.

Delimitations of the Problem

Corporations are in constant competition with each other and sharing analytical methodology is perceived as detrimental to the business advantage the research provides. Therefore, there are privacy considerations with sharing sales data. Many details remain proprietary because the retail company this study explores considers them too sensitive to share. This company has created a model for calculating Remaining Opportunity (RO) and another for determining trade area. Applying both of these models to the sales data helps the company to discover areas with the highest potential for future sales. This study will focus primarily on using the results derived from the models rather than on the intricacies of the models themselves.

GIS technologies are ideal for use in the validation of geographic analysis. In the example presented by Applebaum (1966), potential sales are calculated and then mapped. Later, actual data are collected and used to validate the first predictions. These data are then used to improve the model. After several rounds of modeling and validating, the model is found to be highly predictive. Similar methods were used to determine this company's RO model.

The trade area of a retail location can be calculated in many different ways. The simplest method is to use an existing geographic division, zip code for example. As noted in Tayman and Pol (1995), these types of divisions are frequently too large and diverse. This can lead to over simplification or even misrepresentation of the characteristics of the trade area. When researchers create geographic divisions from point locations, the two most common are radius and travel distance. Radius is simply drawing a circle of a chosen radius around a point and using the area within the circle as the trade area. For travel distance, the time it would take a potential consumer to reach a retail location is taken into consideration. This means that a site near a highway would have a larger trade area than one on a back road. The study will describe this company's hybrid approach to determining trade area.

Concern over keeping data and models proprietary is a major hurdle for researchers interested in using GIS tools in the business environment. While the field of business demography is relative new, the desire to keep business data confidential is not. According to Anderson and Seltzer (2009), the need for confidentiality has existed for at least four decades. They further explain this need is as strong or stronger today. Large corporations have performed studies similar to the one this research proposes; unfortunately, the results are not published. There is limited overlap between the personnel and thinking of business analysts and academia. This situation leads to each corporation creating their own methodologies. It also means that these methods are typically not scientifically validated, simply proven effective by the success of their predictions. There are few studies to base new studies upon, but this also means there is opportunity in the silence on this topic. This study hopes to help establish the academic roots to this type of analysis. The insights brought out by the peer review process will help to mature the discipline which will benefit both scholar and businessperson.

Study Area

The extent of the study area (Figure 1) was the seven counties surrounding downtown Chicago: Cook, DuPage, Kane, Kendall,

Figure 1. Study area.

Lake, McHenry and Will counties. These counties were ideal for three main reasons. First, they represent the majority of the population of the Chicago metro area. These areas contain over 2.5 million people and this company has a low market share. Second, they represent primarily an urban environment. This is important for this study, because the greater size of the study area, the greater the diversity of the demographic it covers. A model that is predictive for urban Chicago may fail to be predictive if it is taken too far outside the city into a more rural/secondary market. A separate analysis could be conducted for rural Illinois, but this study sought to focus on an area of greater impact for the business. And third, the greater the area of a study, the greater the effort required to analyze. By choosing this study area, the size of the area and the time available were balanced.

Methods

Before describing the methods, there are a few key terms requiring explanation. The company described by this study has hired a contact company to do some of the modeling. The modeling service has helped this company by providing demographic and retail databases as well as creating models, some of which are utilized by this study. A prospect is a location that has been targeted as a potential store to be selected. Prospects may be identified in many ways. In some cases a scout will physically drive by a location or in some cases all recently opened competitor locations will be added systematically. There are many, many prospects, and sorting through which might actually be appropriate for this company is a daunting task. For this study, a trade area is the area of land around a retail location that customers are expected to be drawn from. The trade areas have been calculated using a model developed by the contact company that considers many variables, such as proximity to highways and population density.

Finally, the two models this study relies on most heavily are RO and Score. Remaining Opportunity is a simple model created by this company which is used for estimating the remaining dollars in a market. It uses this company's expectation of what market share should be and assumptions about target population and expected sales per client to estimate the potential sales. Then actual sales are subtracted leaving an estimate of the sales that remain to be captured by new locations. Score is a predictive model

created by the contact company which estimates the sales an individual location might generate if it was converted to this company. Essentially RO gives an idea of the potential of the neighborhood and the Score allows for prioritization between specific locations.

Steps of Analysis

After selecting the study area, the next step was to create a 1 mile by 1 mile grid overlaying the study area (Figure 2). The grid was created in proprietary GIS software and then loaded into Esri's ArcMap. One challenge with this step is that Chicago is a coastal city.

Since the grid was created by drawing a shape approximately around the study area, many grid squares fell entirely in the water. In order to include as many grid squares as possible and to capture the economic influence of downtown being situated along the water, each square that was entirely in the water was deleted, while cells only partially in the water were included in the study. This left 4339 grid squares which resulted in a final study area of 4339 square miles.

Esri's ArcMap was used to find the center of each grid square (Figure 3). Using this technique, this method was able to evaluate the entire study area both thoroughly and systematically. Some points fell in undesirable locations, such as in a pond or on top of a government building; however, this study was concerned with the region that the point is in, not the point precisely, so that was not scrutinized.

The hypothetical prospect points were loaded back into the contract company's proprietary software and the trade area was calculated for each point using their model as if each were an actual prospect. The model calculated a trade distance from each point to use as a radius. This is the distance a consumer can be expected to travel to a store.

Figure 2. One mile square grid cells.

The circle created by this radius was then used to create the trade area by including all the zip codes with centroids within that radius. The trade areas are thusly uneven circles of zip codes around each point. Figure 4 shows one inner city prospect and a second prospect at the edge of the study area, in a more rural location. The population density and access to freeways hugely impacts the size of the trade areas.

The next step in the analysis was to link the trade areas for each point back to the grid squares and then sort by size of the radius of the trade area. In Figure 5, red cells represent areas with the smallest trade areas, and then the scale progresses to green, which represents the largest trade areas.

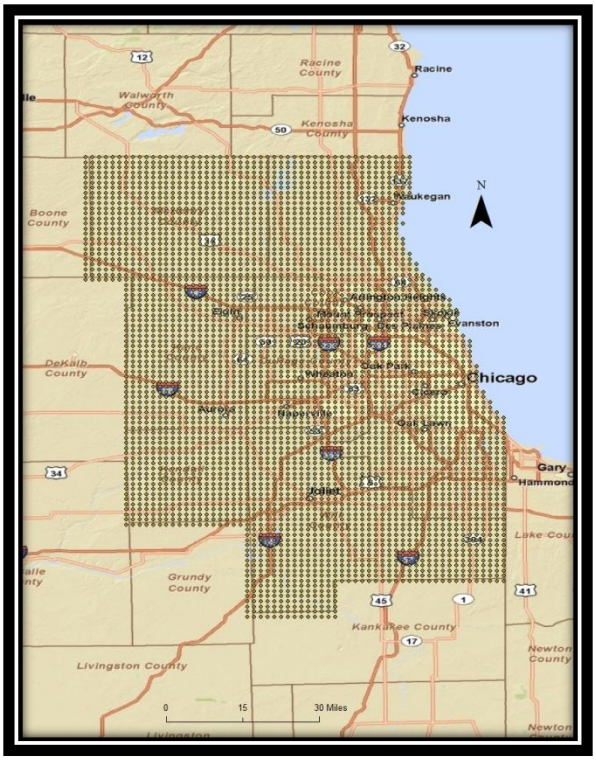

Figure 3. Points representing the center of each grid cell.


Figure 4. Trade area examples.

The trade areas are in general smaller near downtown and grow larger the further they are from the downtown area. However, the size of the trade area does not exactly radiate away from downtown. Some variation of trade area size, especially within the 4-8 mile (orange) area are due to variation at the neighborhood level. The single green square in the downtown area covers a very busy section of downtown Chicago that draws consumers from unusually long distances for access to prestigious shops.

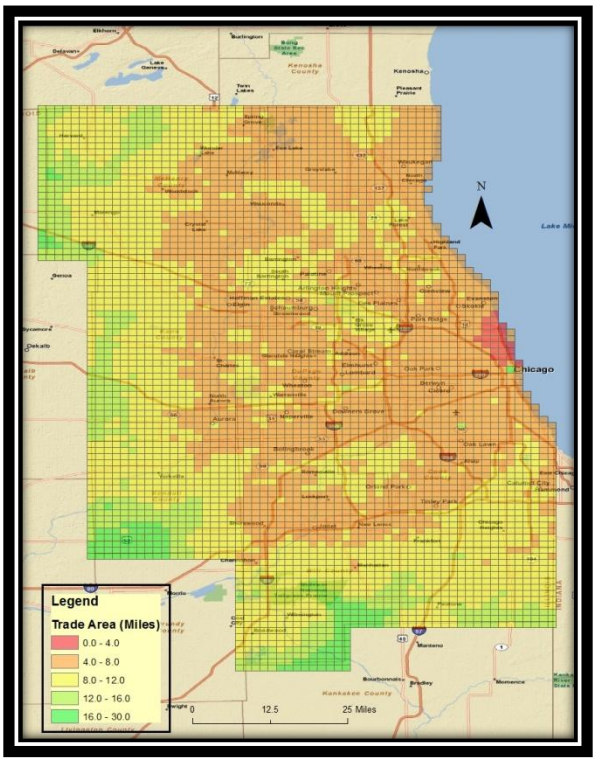

Figure 5. Radius of trade area of each hypothetical prospect.

The final step was to use trade areas to determine the RO for each hypothetical prospect store point, but before that, this study will examine a couple of the factors that contribute to calculating RO. First, Figure 6 shows the median income of each trade area, represented using the grid squares. The areas with the lowest median income, shown in red, are also the most urban. The median income increases as distance from downtown increases, into the suburbs, and then decreases again at the edges of the study area in areas that are very rural.

Figure 7 shows a map of the population density of this company's target population. Essentially these are woman in a particular (proprietary) age range. The green shows the highest

population of this group, yellow is in the middle and red is the lowest population.

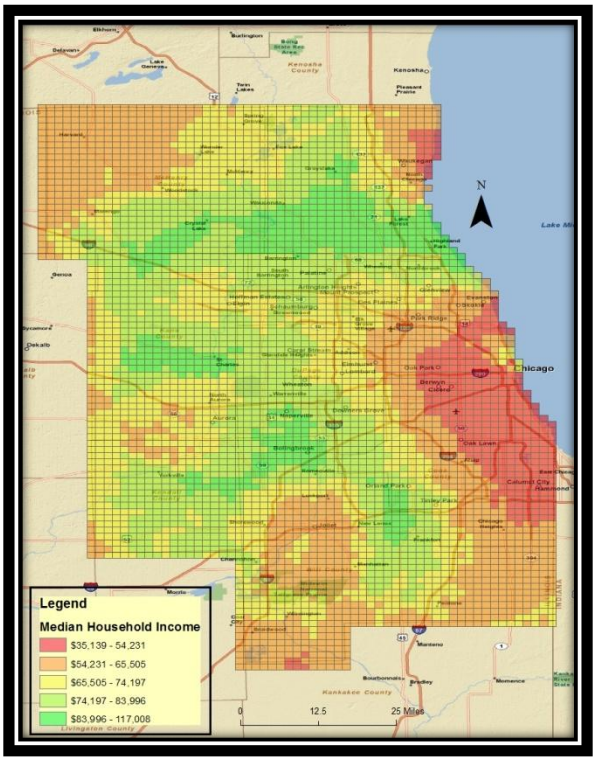

Figure 6. Median household income.

Income level was not used to calculate the target population because income level is relative to location. Consider the status of someone making $75,000 in Minneapolis, MN versus that same individual living in Los Angeles, CA. It is not possible to create one salary range that will capture the target population in all markets.

Moreover, Figure 7 reveals a strong correlation between the density of this company's target population and the overall density of the general population. The highest population is near downtown with population density decreasing outward from there. This suggests this company might benefit from narrowing its definition of what characteristics the target population should have. However, for this study this company's current standard of the target population was used.

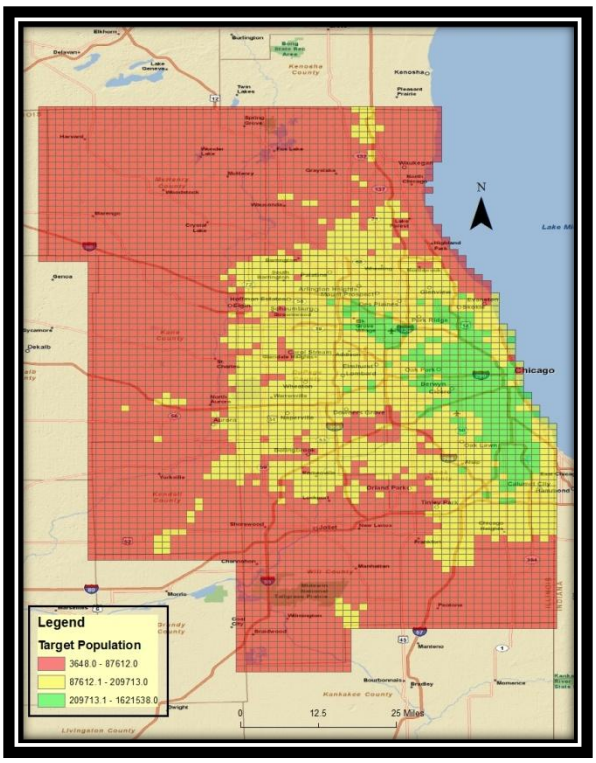

Figure 7. Target population.

In addition to qualifying who this company's target market is, this company also determines a conservative national target for market share. This target is re-evaluated annually and is increased or decreased regionally based both on the median income of the area and also the national median income. The red areas on the map represent the areas where this company's market share is highest. These are areas with less room to add new locations, because this company is already near their target for how much of the market they should be capturing. Conversely, the green grid cells denote areas with low market share and represent regions with plenty of room in the market for this company to grow its share.

Results and Discussion

Figure 8 shows the RO for each grid square in the study area. Remember, the trade areas were calculated by treating the

center of each grid cell as a prospect. The RO of each hypothetical prospect was calculated using this company's

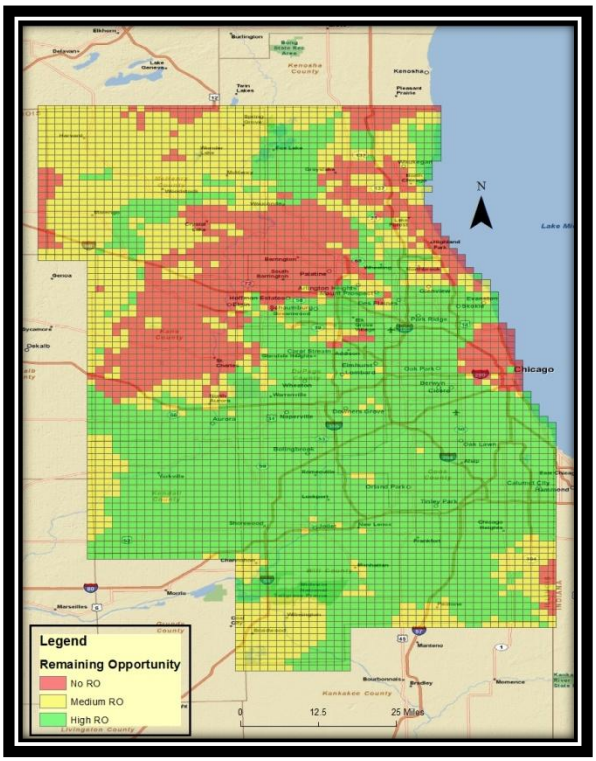

Figure 8. Remaining opportunity.

model for the expected total potential sales for the calculated trade area minus the sales of all the current retail locations within that trade area. The potential sales calculation takes many factors into account including population density of the targeted demographic and this company's market share. In addition, the calculation includes assumptions about how often people visit retail locations and how much they spend while they are there. The factors used in calculating Remaining Opportunity are intuitive, which is one of the strong points of the current company model; it is easy to explain and understand. However, there are also some drawbacks.

Remaining Opportunity is not a predictive but a descriptive model that is a good measure of the magnitude of sales dollars that are available to be captured in each market. The model can be thought of as a screen for good neighborhoods as opposed to an exact indicator of where stores should be located. In addition, the model is limited by the quality of the demographic estimates and their residential nature – that is, they are estimates for people living in certain areas, not for tourists, commuters or most college students.

It is also important to note that while the trade areas of the hypothetical prospects can overlap significantly, each prospect store and the related grid cell were coded based only on the RO of the point within that individual grid square. To clarify, each trade area was created as if the prospect store was the only prospect store, the potential sales of the other prospects stores do not contribute or interfere.

The second model this study used was one that calculated the Score, which is a proprietary model used to predict the sales of a target prospect stores (real or imaginary). The contract company modeled these data extensively and has developed several scores depending on the type of output that is desirable. For this study, score 3 which is a unique and proprietary combination of factors was used. It is of key importance to know that all the models are based on factors in the contract company's database of demographic and other data that show correlation to high sales. That is, the relationships are predictive, but not necessarily causational. Some of the factors used in their model are proximity to shopping centers and what types of retailers are nearby, including the locations of current retail locations.

Figure 9 shows the Score of each grid cell when the center of each square is entered as a prospect store. Large

shopping malls which are represented by red circles increase the diameter of the gross leasable area (GLA) and this area increases as the mall size increases. Clearly the presence of a large shopping mall is strongly correlated with

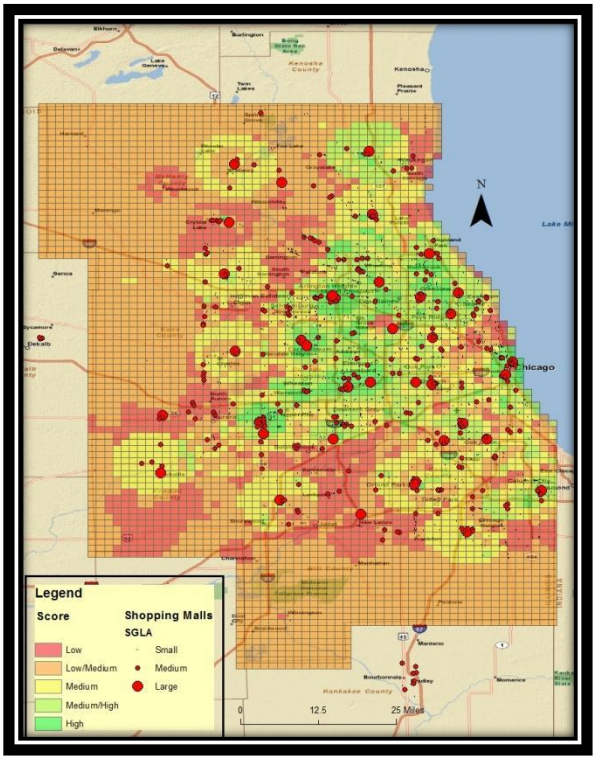

Figure 9. Score for hypothetical prospect store in each grid square.

an increase in the Score in the vicinity, which means that the increasing GLA is correlated with increased sales in nearby stores.

One point to note is that unlike RO, the Score is not concerned with market penetration. It is very likely that some of the areas that show high Score would not make good locations for new stores because the neighborhood is already saturated. The best use of this model is as a method for comparing specific prospect stores in a neighborhood already known to be desirable.

After using RO and Scores to model the study area, this study evaluated some stores that are real life prospects. These prospects came from a database maintained by this company. The salespeople who maintain this database have already eliminated about half the locations in Chicago that would not be suitable for this company. The 5600 locations in the database were exported. Figure 10 offers a good visualization of how challenging it is for scouts to select which prospect stores to target for conversion.

The result of using the grids with RO to narrow down the prospects by performing a spatial join between the prospect stores and the RO grid is shown in Figure 10. This join linked each prospect store to the grid it was located within which enabled the study to determine if its location did or did not have RO. Green prospect stores were located in grids that had high RO, yellow were in areas with medium RO, and prospect stores in red were located in areas with no RO.

In addition to linking the RO of the area, the Score was also calculated for each real prospect store. This was not determined using the grid but rather using the actual prospect locations.

For this analysis to be useful to this company's scouts, the list of prospects needed to be reduced and prioritized. The data were exported into Microsoft Excel for additional analysis. One extraneous question these data were able to address was whether RO and Score were correlated. The answer was, yes, but negatively. This seems counter intuitive, but can possibly be explained by the fact that the areas with the best Score already have locations which would reduce RO due to the market being saturated.

Once these data were in Excel, the next step was to filter the prospects by those in areas that still had market share,

which were those with RO. The prospects with no RO were removed. That left 4200 of the original 5600, which is progress, but the list still needed to be prioritized.

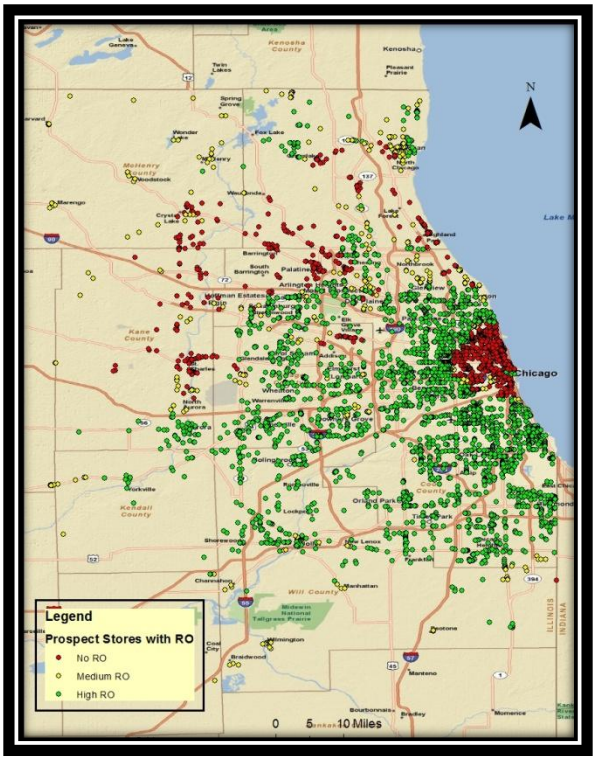

Figure 10. Remaining opportunity of real prospect stores.

The final step was to further reduce the number of prospects by using the Scores as a second filter. The stores were sorted by score and entries that did not meet a specific minimum value were deleted. This decreased the total number of prospects down to approximately 3400. This study provided this company's salespeople and scouts with a list of 3400 prospect that are locations with the highest RO and at least the minimum Score. This list was also sorted by highest Score so efforts could be efficiently prioritized.

Conclusions

Corporations who wish to expand their sales in today's tight economic conditions require smarter, more efficient ways of doing business.

This study showed that using GIS to locate optimal retail locations is one such method. Globally, GIS is being used to identify market areas and predict customer demographics. It is being used to target new retail sites and identify sites that should be expanded. The possibilities are multiplying every day and any business that wants to be ready for the future needs to incorporate the use of GIS analytical methodology.

Remaining Opportunity and Score gave excellent insight into the suitability of prospect stores without the work of contacting or visiting. This analysis showed that it is possible to prioritize prospect stores by how successful they are likely to be if converted to this company's location. This study should make the Chicago sales team more efficient and is an example case of what is possible.

As suggested in Vlachopoulou, Silleos, and Manthou (2001), GIS findings can be combined with less technically demanding user interface software in order to assist in site selection decision making. While this study has shown how great the potential of this type of analysis is, it has also shown that the need for human work in sorting through prospects cannot be eliminated. Perhaps it is a more expansive truth that analytical methods for categorizing data are only useful for a certain level of detail.

This analysis, however, could be expanded beyond this study. The same procedure could be performed beyond Chicago and even beyond this company. It could be modified to work for any company that has retail locations. It could be applied across the United States and to any country with reliable census data. The analysis could also be refined to be more precise. Perhaps soon, this company will

be able to look at where successful Chicago new stores are located in comparison to where this study predicts they should be. Using RO and Scores to reduce the number of prospect stores was successful but the result still included many more stores than a scout could actually visit. Conceivably the list could be filtered again by median household income or perhaps it could use a closer examination of how much cannibalism might occur if a new store were to open too close to an existing store.

Many other similar companies and likewise other industries need to select new retail locations. This method could be modified to work in essentially any retail site selection analysis. A company could use this study to create similar methods of their own.

Acknowledgements

Special thanks to Greta Bernatz who provided invaluable insights as academic advisor at Saint Mary's University of Minnesota. Without her support, this project would not have been possible.

References

Anderson, M., and Seltzer, W. 2009. Federal Statistical Confidentiality and Business Data: Twentieth Century Challenges and Continuing Issues. *The Journal of Privacy and Confidentiality*, 1(1), 7-52.

Applebaum, W. 1966. Methods for Determining Store Trade Areas, Market Penetration, and Potential Sales, *Journal of Market Research*, 3, 127-41.

Hernández, T., and Bennison, D. 2000. The Art and Science of Retail Location Decisions. *International Journal of Retail and Distribution Management*, 28(8), 357-367.

Kimes, S.E., and Fitzsimmons, J.A. 1990. Selecting Profitable Hotel Sites at La Quinta Motor Inns. Interfaces, 20(2), 12-20.

Rogers, D. 1987. Shop Location Analysis. McFadyen, E. (Ed.), The Changing Face of British Retailing, Newman Books, London.

Simkin, L.P., Doyle, P., and Saunders, J. 1985. How Retailers Put Site Location Techniques Into Operation: An Assessment of Major Multiples' Practice. *International Journal of Retail and Distribution Management*, 13(3), 21-26.

Tayman, J., and Pol, L. 1995. Retail Site Selection and Geographic Information Systems. *Journal of Applied Business Research*, 11(2), 46-54.

Vlachopoulou, M., Silleos, G. and Manthou, V. 2001. Geographic Information Systems in Warehouse Site Selection Decisions. *International Journal of Production Economics*, 71, 205–212.