Utilizing Geographic Information Systems to Locate Target Markets in the Retail Banking Sector

Todd H. Zielsdorf
Department of Resource Analysis, Saint Mary's University of Minnesota, Winona, MN 55987

Keywords: Geographic Information Systems (GIS), Marketing Analysis, Target Marketing, Spatial Analyst, Demographics, Marketing, Banking Products, Geo-demographic

Abstract

GIS was used to analyze potential markets within the seven counties that make up the Twin Cities metropolitan area. A raster-based, weighted model along with demographic data were used to find the most ideal location(s) for marketing to potential and current customers, while simultaneously isolating the most optimal location to focus first marketing efforts. It was determined that high-earning families with children offer the most potential for expanding, or attracting customers to a multitude of banking products that include savings bonds, retirement accounts, college savings funds, regular savings accounts and certificates of deposit. This model can be adjusted and used in the future to find secondary locations for additional marketing efforts.

Introduction

Demographic data including median household income, percentage change in median income, and number of persons in the household, provides an ideal opportunity to analyze economic characteristics by census block and census block groups. Multiple census blocks make up a census block group (Figure 1).

Blocks and block groups define demographic homogenous areas created from socioeconomic groupings of like neighborhoods (www.census.gov). By using census blocks and block groups, a bank marketer is able to break down and isolate areas into smaller geographic sections that provide the most ideal locations to find bank customers. Since a customer is the source of revenue for a bank, using GIS to find this focus area is crucial. Financial service organizations "cherry pick" their customers for high value, low risk financial products, and by using GIS, banks can target these customers more efficiently (Talbot, 1998).

Figure 1. Alignment of census blocks within a census block group.

Zielsdorf, Todd. 2008. Utilizing Geographic Information Systems to Locate and Target Markets in the Retail Banking Sector. Volume 10. Papers in Resource Analysis. 9 pp. Saint Mary's University of Minnesota Central Services Press. Winona, MN. Retrieved (date) from http://www.gis.smumn.edu

In the banking sector, emphasis on up-selling, or cross-selling products to current customers is more economically feasible than relocating a bank to where customers exist. Up-selling involves informing a customer of products that are similar to what they currently have, but might cost a bit more, while cross-selling involves selling to customers similar, or complementary products. Prime location sites are increasingly hard to find, with good locations being expensive (Talbot, 1998). The goal of this analysis was to find these customers located in the Twin Cities (Minneapolis/St. Paul) metropolitan area of Minnesota. This area consists of seven counties, including Anoka, Washington, Hennepin, Ramsey, Carver, Scott, and Dakota. High-income families with children provide a unique opportunity for targeting additional or new banking products. Higher income, according to the Consumer Expenditure Survey from US Department of Labor Bureau of Statistics, is yearly income over $70,000. High-income parents are more likely to purchase savings bonds, college savings accounts, mutual funds and certificates of deposit for their children, while pursuing retirement accounts and home mortgages for themselves. This broad demographic group between the ages of 18-39, consisting of high earners with children, was the target market for this project. The goal of this study was to spatially locate people fitting the demographic profile created by this study.

Methods

Technology

Software used in the production of this project were ArcGIS 8.3 and ArcGIS 9.2, both developed by Environmental Systems Research Institute (ESRI). The Spatial Analyst was used to create the multitude of raster layers used in the analysis.

Data Collection and Preparation

Shapefiles for census blocks, census block groups, roads, county boundaries, and municipal boundaries for the Twin Cities metro area were all downloaded by connecting to the MetroGIS server in ArcCatalog (datafinder.org). US Census tabular demographic data including SF1 (Summary File 1) and SF3 (Summary File 3) were downloaded from the US Census Bureau's website (http//factfinder.census.gov). SF1 data is compiled by the census block, which is the smallest and most ideal dataset to use. While this dataset would offer the most accurate separation of characteristics, it does not include economic data that is needed for this project. SF3 data is compiled by census block group, generally consisting of populations ranging between 300 – 3000 persons, or 250 – 550 housing units, and contains the economic and income data needed. SF1 and SF3 tabular census data were joined with census block, and block group shapefiles through a join function. This resulted in all census blocks and block groups for the metro area containing the tabular data needed for the project, including; median household income, number of renters, average household size, percentage change in median income, population, along with specific population figures by income levels, and age groupings. Since median income eliminates any extraneously high or low values, median income was used versus using average income figures because

very high or low numbers could skew the results. For example, if a census block contained a few families with extremely high income and average income was used, those high income families would result in an artificially inflated income level for the whole census block and producing skewed results.

Initial Analysis and Data Manipulation

The first analysis was conducted to locate high-income areas by executing a spatial query of yearly income greater than $200,000 to produce a dataset for verifying against the weighted suitability model. This was done in order to generalize the data for future verification and to give an overview of areas containing the highest incomes (Figure 2) to compare to the output of running the weighted suitability model. Findings can then be used to identify potential errors during analysis.

Raw Data before Analysis

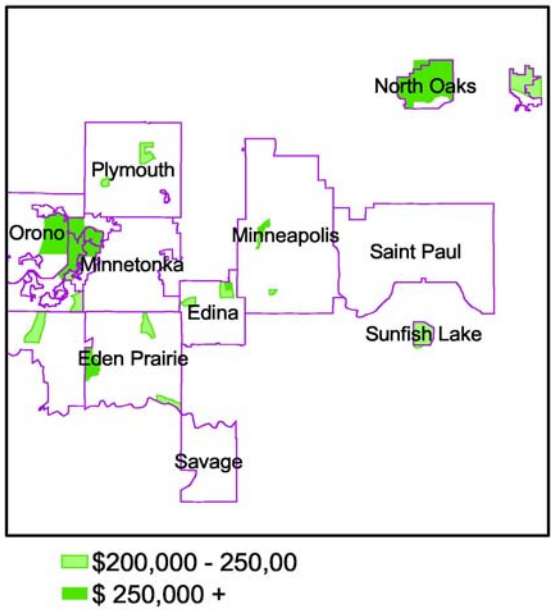

Figure 2. Initial analysis of raw data.

The next task was to prepare and visually inspect the following demographic variables, which were to be used for this project. How these variables were subsetted and used in analysis are defined in the "Model Analysis" section.

-Median Household Income
-Percentage Change in Median Income
-Population of Census Block Group
-Household Size
-Renters per Census Block Group

Median household income and percentage change in median household income were used as the best indicators of wealth in a block group. Household size was used to indicate the number of children present in the household, along with population numbers to indicate highest priority of marketing probabilities. The higher an income, the higher value it was given towards potential marketing opportunities. Likewise, the highest household numbers were the highest values to be used in the model to represent the likelihood of children. Renters by census block, were used to identify potential homebuyers. By creating thematic layers and classifying features with the above variables, visual analysis was conducted to compare these results with results obtained from the weighted suitability model. Results from the initial analysis were as follows: there were 22 block group locations in the metro area with a median household income of $200,000 and above, with a total population of 23,155 and a median household income of $235,999. Upon viewing the data, it was determined that renters per block group data had too many null values to be valid for this project (Figure 3). This is likely the

result of renters being difficult to count, or track based on the propensity to relocate.

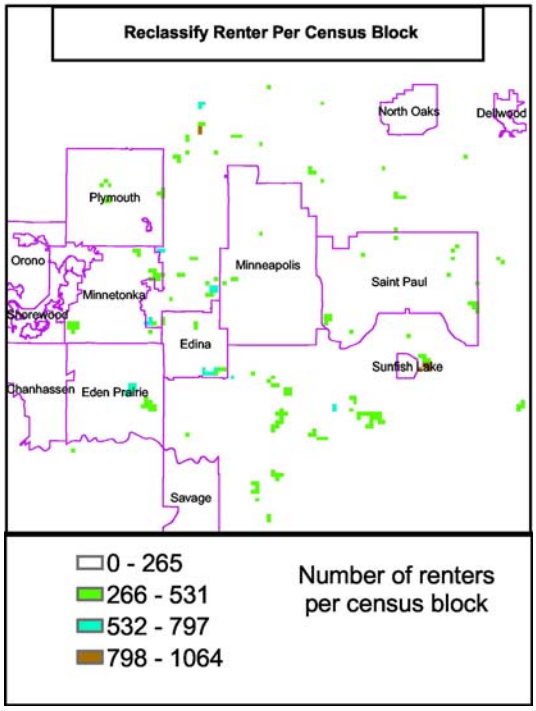

Figure 3. Renters per census block group with many null values.

Model Analysis

Raster data and raster calculator were used to conduct weighted model overlay calculations. Raster data for all variables were generated and reclassified into 4 categories. The median household income and average household size used 4 categories with values of 0 to 3. Percentage change in median household income, and population used 4 categories with values of 1 to 4. Values of 0 were not calculated. Values of 4 were the highest and, the most optimal value in the analysis. Excluding 0, excluded household size of 2 and below, indicating no children present. See the section "Sources of Error" at the end of the paper for more information on considering single parents with children as a family. At this point, raster grids were created using "equal area" classification of the base variables, median household income, percentage change in median income, population per block group, and number in household. Equal area classification is best suited to show the difference between high and low values for each class; it can emphasize the location of the few blocks with the very highest median income (Mitchell, 1999). The values and figures of these raster calculations are subsequently described below.

Median household income was converted to a grid and reclassified. A zero value was used in this raster to eliminate income levels of persons earning below $100,000. This grid was used to target high earning individuals (Figure 4).

$0 – 99,999 = 0
$100,000 - 149,999 = 1
$150,000 - 199,999 = 2
$200,000 – 291,862 = 3

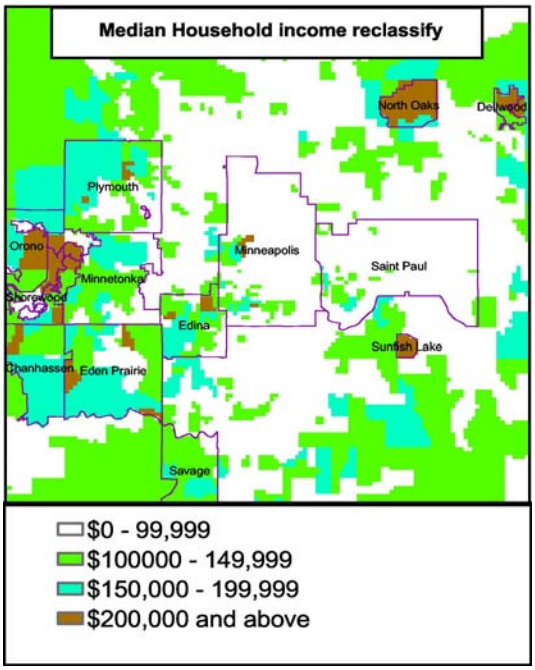

Figure 4. Reclassification of median household income.

An average household size grid was generated and reclassified using a 0 value in the first category of 0 – 2 persons to eliminate singles and couples. The household size classification was divided into the following intervals (Figure 5).

0 – 2 = 0 Persons
2.01 – 2.99 = 1
3.00 – 3.99 = 2
4+ = 3

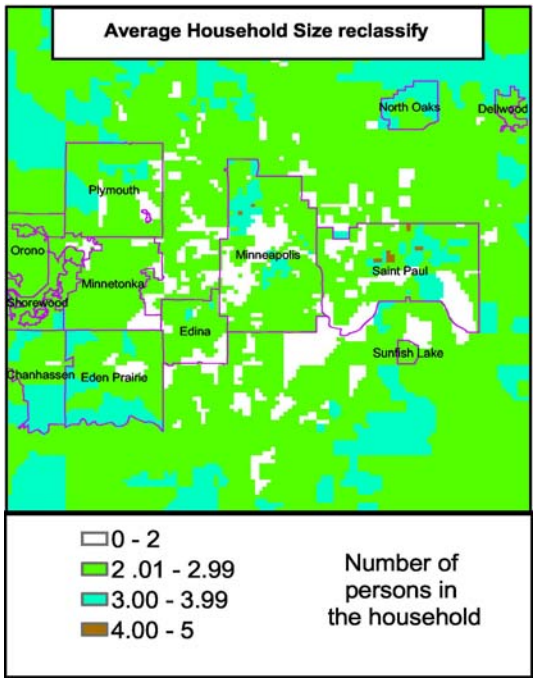

Figure 5. Reclassification of household size grid.

A percentage change in median household income grid was created and reclassified according to the following intervals (Figure 6).

0% – 13.99 % = 1
14% – 27.99 % = 2
28% – 41.99 % = 3
42% – 56 % = 4

A 2004 population grid was produced and reclassified according to the following intervals (Figure 7).

0 – 2103.25 = 1
2103.25 - 4206.50 = 2
4206.50 - 6309.75 = 3
6309.75 - 8413.00 = 4

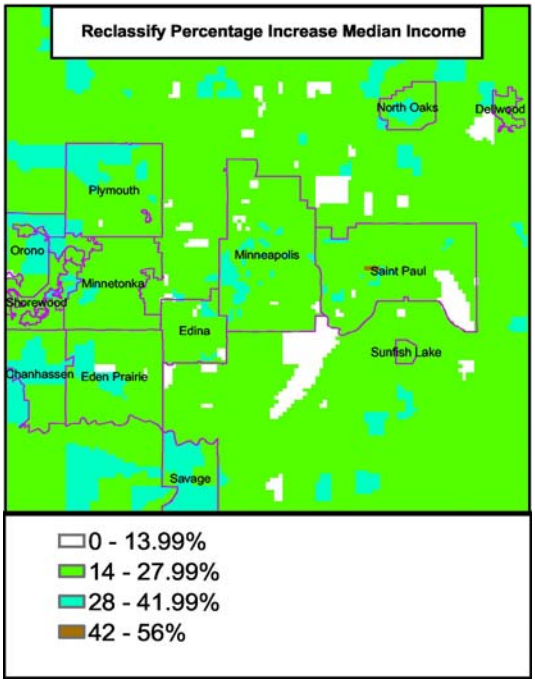

Figure 6. Reclassification of grid containing percentage change in median household income.

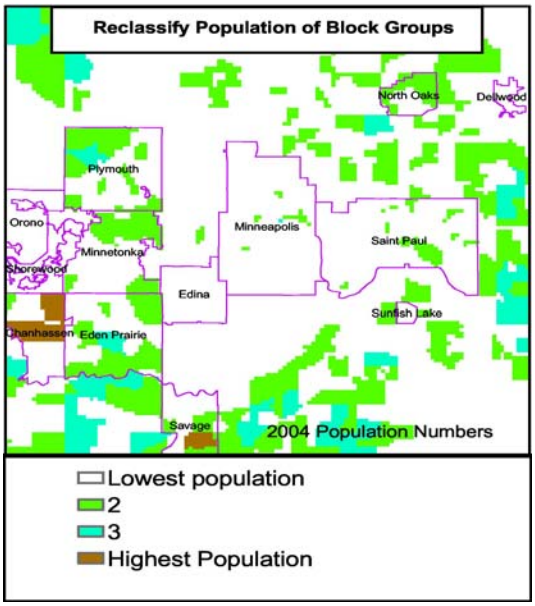

Figure 7. Reclassification of population data into 4 values.

The previous 4 variables were then used to create a weighted model grid. Through the use of the raster calculator, a base marketing grid was created using the following data and percentages.

Median Household Income	50%
% Change Median H.H. Income	25%
Population per Block Group	25%

This weighted model was created based on the idea that high earning individuals provide the most revenue opportunities for banks to target new or additional products. Wealthy areas are more likely to respond to direct marketing of financial services versus low-income areas. Median income was the most important value based on reliability of income level being a determinate of propensity to take out financial services (Berry and Longley, 2005).

Population and percentage change in income were used at 25% each to, first indicate any recent income changes to specific block groups since the 1990 census was conducted, and secondly, to base the data on population numbers which allocated higher value to income levels (Figure 8).

This base marketing grid was then reclassified following the same format of using 4 classes to keep values equivalent for use in raster calculator overlays. For display purposes, grids of 10, 15, and 20 classes were also created to observe how changing class size affected findings. Grids of 15 and 20 classes produced one ideal overall marketing location, Figure 9.

This base marketing 4-class grid was then used in the raster calculator to analyze a household size grid in relation to overall marketing potential (note, the renters grid analysis was eliminated due to data having null values). This step involved creating a grid with the raster calculator using the overall marketing

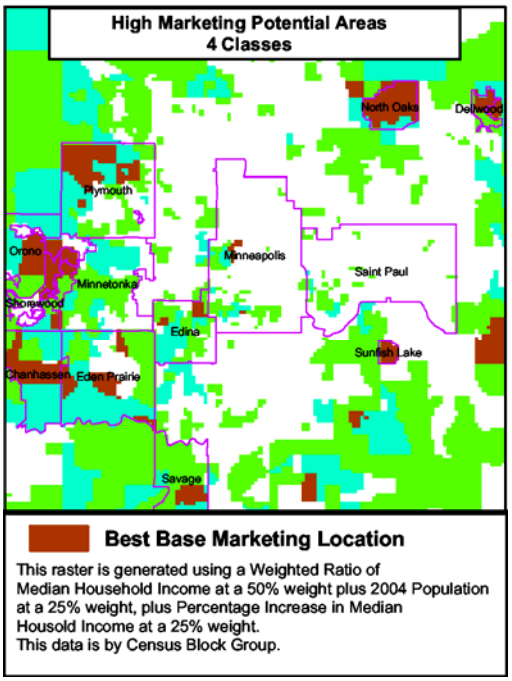

Figure 8. Base overall marketing grid generated to 4 values.

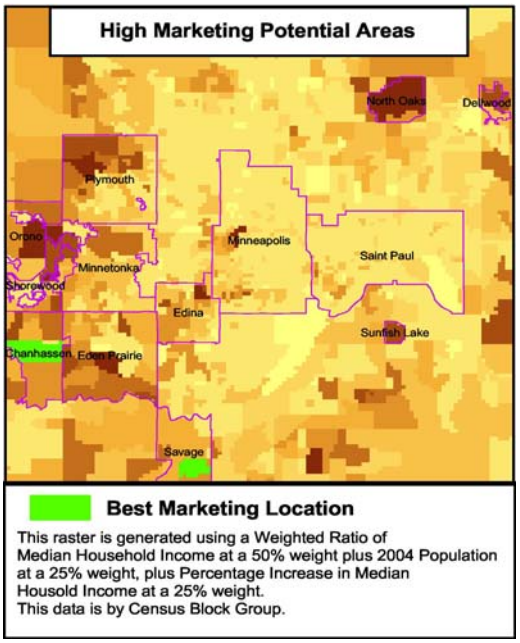

Figure 9. Multiple opportunities, base marketing grid isolated by equal area classes to 2 locations.

grid at 25% weight, combined with the household size grid at 75%. This grid produced a dataset indicating the best possible location to begin target marketing to high-income families with children.

In reiteration, this grid was classified as equal area by testing an increasing number of classes until a graduated grid was developed to narrowly locate one, or possibly two locations (Figure 10).

The target-marketing grid located the highest income families with the highest number of children. This raster analysis indicated a block group in Savage, Minnesota as the ideal location for focused marketing efforts. For further analysis, both the main marketing grid (Figure 9) and the focus grid (Figure 10) can be used to locate secondary areas by changing the number of classes.

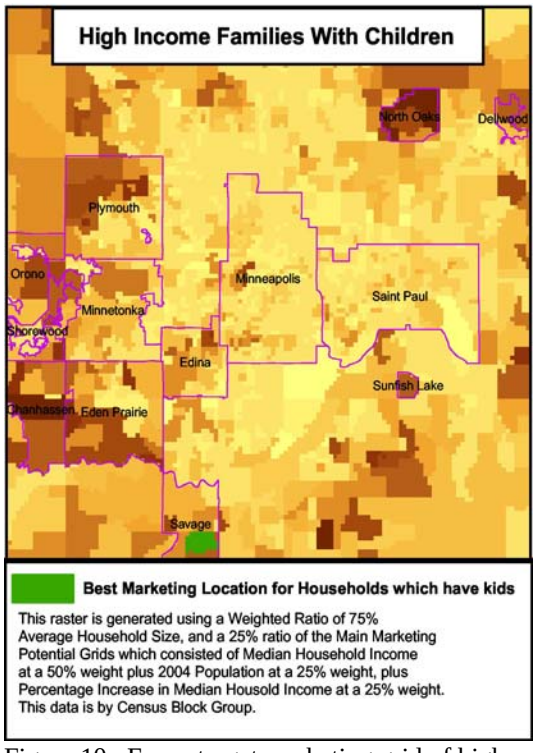

Figure 10. Focus target marketing grid of high earning families with kids, isolated by classes down to one area.

Results and Discussion

The target market isolated in this project will provide a great "first tier" marketing effort. By using these results along with a banks internal customer data, total market penetration and current market penetration can be determined. Using GIS analysis to market in the ideal and most probable locations first can save time and money.

While this project used data that can be obtained relatively easy and freely, it was GIS that created a useful predictive model to spatially map patterns of high potential areas to target banking products. An example of this can be seen in comparing the spatial query of raw data, with that of the overall marketing analysis grid, and the focus area analysis grid (Figures 2 and 9, respectively). Raw data only contained areas with median household incomes over $200,000, while using a weighted model led to the focus area in Savage, Minnesota that included high income persons along with high population, and higher number of children. To further enhance a project of this nature, specific demographic data could be purchased from any of a number of companies that specialize in data development. For the scope of this project, free datasets worked very well to indicate initial targeted marketing towards high earning families. Also, to locate secondary marketing opportunities, any of the models could be adjusted by either adding more classes to the 4 variables of median household income, percentage change in median household income, household size, and/or population. This project was well suited as an example in finding the most ideal location to begin a first phase target marketing effort, and as time progresses, changing the

reclassification to locate the next best areas to pursue. Figures 11, 12, and 13 illustrate descriptive statistics for the specific target area and were derived from block and block group data. This information can also be used to advertise particular products to this specific location in the form of billboard, radio and site specific advertising, along with census block group mailings.

As mentioned earlier, high income persons in the 18 – 39 demographic age group are most likely to be starting careers, and/or are recent graduates, or new professionals to an area, purchasing a home or retirement account for themselves and with children, likely investing in their children's future in the form of CD's, regular savings, mutual funds, and college savings accounts. Consumers within neighborhoods influence the pattern of consumption of each other, resulting in similar patterns of product consumption within that neighborhood.

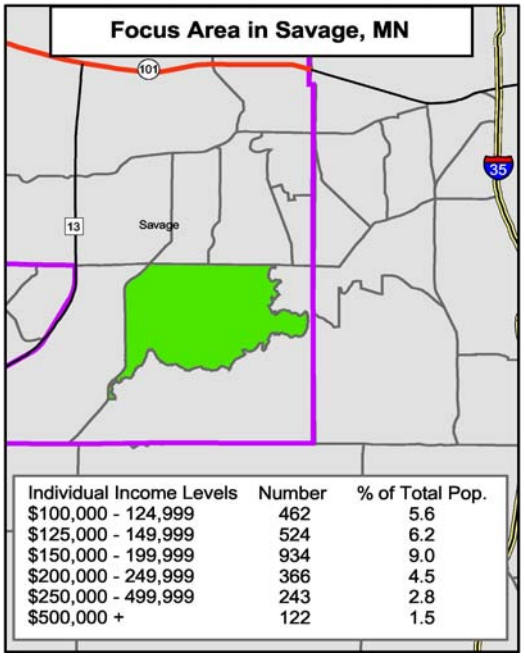

Figure 11. Focus area census block group, including income level statistics.

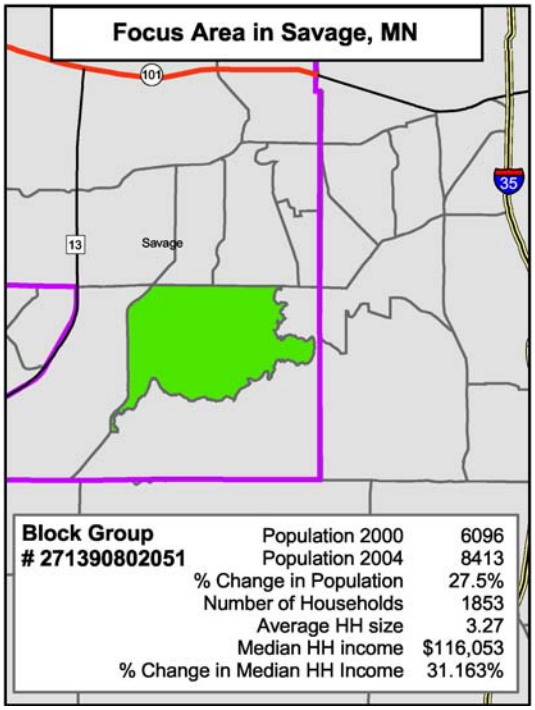

Figure 12. Block group demographic statistics.

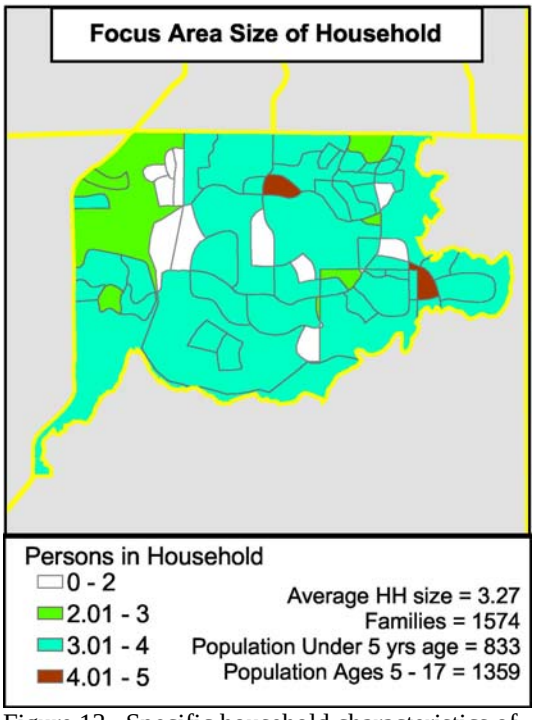

Figure 13. Specific household characteristics of focus area block group.

Sources of Error

In considering use of the average household size variable, and allocating a

0 value to the 0-2 persons grouping, resulted in eliminating any single parents with a child. This was not considered to be a serious source of error however as in most instances, a single parent family would most likely be devoting significant energies and finances towards family needs and not be identified as an ideal customer anyway.

Acknowledgements

I would like to thank all my colleagues at Saint Mary's University for the opportunity to learn and progress from the many avenues of education and experience offered. I would also like to thank my professors David McConville, Patrick Thorsell, and especially John Ebert for continually giving guidance and support throughout this project.

References

Berry, S.J. and Longley P. 2005. Assessing the usefulness of store card data in direct sales of financial services. *Journal of Retailing and Consumer Services.* 407-417.

Mitchell, Andy. 1999. The ESRI Guide to GIS Analysis. Volume 1: Geographic Patterns and Relationships. 48-49. Environmental Systems Research Institute, Inc.

Talbot, Chris. 1998. Geographic Information System Applications in the Retail Banking Sector. *OXIRIM.* Retrieved: November 11, 2005 from EBSCO*host* database.

US Department of Labor Bureau of Statistics. 2008. Technical Documentation. http://www.bls.gov