POPULATION DISAGGREGATION FOR TRADE AREA DELINEATION IN RETAIL REAL ESTATE SITE ANALYSIS

by

Alfredo David Cisneros

A Thesis Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

May 2015

DEDICATION

I dedicate this document to my parents for their constant support, my family and in laws, and most importantly to my wife who spurred me to get it done and agreed to make me the happiest man in the world.

ACKNOWLEDGMENTS

I will be forever grateful to all my professors in the GIST program, and most of all Professor Karen Kemp. Thank you to my committee members Victor Bennett and Katsuhiko Oda. Thank you also to my family and friends, all of whom displayed patience, stuck with me, and all inspired me throughout this process and helped me complete this degree program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAG	Association of American Geographers
AVGHHINC	Average Household Income
AVGNW	Average Net Worth
GIST	Geographic Information Science and Technology
PCI	Per Capita Income
POP	Population
PP	Populated Parcels
SSI	Spatial Sciences Institute
USC	University of Southern California

ABSTRACT

An appropriately sited retail location can turn a business into a veritable cash machine for the owner. Siting a store location has financial implications for store owners, banks, real estate professionals, store employees and company shareholders, all of whom are impacted by the success or failure of a store. Determining catchment population -- the population within a store site's actual or potential trade area -- is essential for good retail site suitability analysis. An accurate calculation of a store's catchment population depends on the method of defining a store's trade area and the accuracy and the precision of population data.

This study explored how concentrating aggregated census population into existing developed residential areas affects the results of trade area analyses likely to be used in retail real estate marketing and decision making. Different methods of defining trade areas were also used to explore how the differing trade area outcomes affected results of analyses used for retail real estate decision making. It also seeks to show how different store sites with different population densities ranging from very dense areas in suburban areas to areas bordering rural areas affect population aggregations.

Results of these analyses showed only small changes in catchment population and demographics when concentrated population areas were used in calculations as opposed to census aggregates. Conversely using different distance measures for trade area creation resulted in large differences in catchment population which should be taken into consideration for analysis and marketing moving forward.

CHAPTER 1: INTRODUCTION

An appropriately sited retail location can turn a business into a veritable cash machine for the owner. Conversely, a poorly chosen site location may result in owners losing all investment in their retail location. For first time business owners this could bankrupt their business and cost them most if not all of their wealth. Siting a store location has financial implications for store owners, banks, real estate professionals, store employees and company shareholders, all of whom are impacted by the success or failure of a store.

When searching for an ideal retail site for a client, most retail real estate brokers can readily list a number of desirable site characteristics that can be used to assess a site's suitability. These might include requirements such as:

- Busy retail area
- Signalized intersection
- Property located on a hard corner
- Average household income within three miles above a given threshold
- Population within three miles above a given threshold
- Daily traffic count on the main road above a specific number of vehicles

All of these characteristics are considered favorable in insuring there is sufficient exposure and market potential for the retail location.

Determining catchment population -- the population within a store site's actual or potential trade area -- is essential for good retail site suitability analysis. Population and associated demographic data are collected by the Census Bureau and to protect individual privacy are reported at various levels of aggregation (i.e. block, block group and census tract). Due to the fact that trade areas and aggregated census units are rarely identical and population is often distributed unevenly within census aggregates, using census data to accurately determine population within a retailer's trade area is problematic.

This research addresses these shortcomings by using land classification and supplementary data from San Diego County to disaggregate biennial census and American Community Survey data in new ways. These disaggregated data were then used to conduct trade area analyses. Finding better methods to disaggregate data may result in more accurate estimates of population distribution and allow for better calculation of catchment population and associated population characteristic statistics.

1.1 Summary of Methodology

This process used three sets of polygon layers in the study area along with census block attributes to disaggregate the census block data. First the areas that were strictly residential were extracted from the land use classification layer. Overlaying this with the parcel layer allowed for the extraction of a "Residential parcels" layer separating residential land use areas into smaller parcel divisions. A third polygon layer of developable land used was used to remove undevelopable areas as designated by the county in order to concentrate the population into "Developed residential parcels". Finally, Census blocks without population were removed to further concentrate population leaving only populated developed residential parcels, the population data source used in the analysis referred to as "Populated parcels".

These new data were used to determine a site's catchment population for detailed drive time trade areas to see if there were differences in the calculated population and related population characteristics than when calculated by the more commonly used manner which employs the census data polygons. Four store locations for suburban and rural cities of San Diego County were chosen to see the effects of the resulting variations in population density on these analyses. Drive Time trade areas for each site for a store of non-specialty goods were calculated at three, five, and seven minute thresholds. Five, ten, and fifteen minute trade area thresholds were calculated for specialty goods store sites. A manual calculation of the same spatial analysis process was conducted to confirm the calculations of resulting catchment population and characteristics were accurate.

Distance measures in miles are more commonly used to create trade areas in real estate marketing and decision making due to fewer data requirements needed to create these trade areas. Trade areas were also created using 3, 5, and 7 mile road network and radial distance measures. Catchment population and demographics were calculated for each of these trade areas using census aggregate and “Populated parcels.” These results were compared with trade areas calculated with drive time minutes created previously to see how these differences could impact marketing and decision making.

1.2. Research Goals

This case study explores how concentrating aggregated census population into existing “Populated Parcels” affects the results of trade area analyses likely to be used in retail real estate decision making. Different methods of defining trade areas were also used to explore how the differing trade area outcomes affected these analyses as well as similar analyses used in retail real estate decision making. It also shows how different store sites with different population densities ranging from very dense areas in suburban areas to areas bordering rural areas affect population aggregations.

1.3. Structure of Thesis

The next chapter reviews related research that has been undertaken previously on determining site suitability for retail sites, methods for disaggregating census population data to determine sales potential, the definition of trade areas and other fundamental themes in this thesis. Details of the methodology outlined above are discussed in Chapter 3. Results of the analyses are discussed in Chapter 4. Conclusions from this study and recommendations for future improvements and studies are discussed in Chapter 5.

CHAPTER 2: BACKGROUND AND LITERATURE REVIEW

Retailers provide products to the public for consumption. In *Intelligent GIS*, Birkin et al. (1996) review how the success of a retailer is dependent on how well they execute on the retailer marketing mix. The retailer's marketing mix is made up of Product, Price, Promotion, and Place commonly referred to as the 4 Ps of marketing. Product refers to the good or service available for consumers. Demand for products is determined by consumer's level of need and product desirability. If there is demand for a product then the consumer considers the second P price, at which the product is available. Price has a negative impact on demand meaning the higher the cost of a product the less likely a consumer is to purchase the product at all or from that particular retailer.

Birkin et al. (1996) explain that if a desirable product is available at a price consumers are willing to pay the retailer relies on the third P, promotion, to inform consumers the product is available and appeal to consumers to purchase from the retailer's location. Promotion can cause a consumer to visit a particular retailer when products or pricing are similar. Promotion is especially important in today's market with many online retailers offering similar or identical products at lower prices. Competing with other retailers and online retailers is difficult but the immediacy of a purchase from a brick and mortar retailer can make a large difference in a consumer's decision. If a retailer is located near consumers and can be reached with minimal difficulty, then consumers are more likely to patronize this location.

Place, the fourth P of the marketing mix, refers to the retailer's location. Distance to a retailer's location has a negative relationship with sales. Hence, the greater the distance the less likely a person will be to purchase a product from that location. Other factors can also influence a consumer's decision to purchase from a store location such as nearby attractions (including stores) as well as other products offered by a retailer at their store. Location also plays a large role in product pricing as real estate costs for all stores are factored into the price of products offered at a location (Birkin et al. 1996).

As location is integral in determining site suitability and pricing, it is crucial to find the right site at the right price for a retail location. Real estate professionals should understand the importance of finding the best site to the business owner's ultimate success. From my experience working with McDonald's corporation to locate sites for their new locations, larger corporations understand the importance of location to their businesses' success and have developed their own departments for finding ideal new sites.

The use of site suitability analysis in retail market analysis has become much more sophisticated and even the smaller companies now use GIS with demographic or traffic count data to make models of varied sophistication to estimate sales potential (Birkin et al. 1996). Sales potential for a retail location is the estimate of retail sales for a specific time frame, usually a one year period for existing stores or multiple years for new stores to allow them to reach profitability. Accurate estimation of sales potential is crucial to the success of a retailer and better techniques for calculating a site's catchment population are crucial to improving these estimates.

A store is dependent on the local population likely to be consumers at that store site. Catchment population is the population within a store's trade area which is more likely to patronize the store location. To determine a store site's suitability for future success, knowing the catchment population's counts as well as demographic and consumer spending information is very important (Birkin et al. 1996).

The accuracy of a calculation of a store's catchment population is dependent on the method of defining a store's trade area and the accuracy and precision of population data. As such, this chapter reviews past and current methods of predicting sales potential for retail sites and the trade area method used in this case study. In particular Esri's Business Analyst trade area tools are discussed including the various trade area creation methods used in this study. This is followed by a review of Census data and issues arising from its use in such analysis. Lastly, a review of areal interpolation methods utilized in the past, and specifically dasymetric mapping techniques similar to those used in this case study are discussed.

2.1. Methods for predicting sales potential

Fenker and Zoota (2000) and Birkin et al. (1996) review intuitive models of predicting sales which can be applied to modeling and GIS approaches for predicting sales potential. Sales potential estimation in the past has followed approaches referred to as analogue, regression, and trade area methods (Fenker and Zoota 2001; Birkin et al. 1996). A summary of these approaches are reviewed in subsequent sections.

2.1.1. Analogue Method

Use of the analogue method involves seasoned real estate professionals and business owners making decisions using the confluence of their experience. They apply known past store location successes and failures to prospective sites determining suitability and potential future success from past sites with analogous characteristics and their "gut feelings". Sites deemed similar to past successful sites are treated as favorable whereas sites similar to past failures are treated unfavorably. This method is subjective because it relies on the judgment of people and a "gut feeling" based on their experience from site visits. Given their subjective nature, these judgments may be clouded by emotion surrounding a past experience and are less objective than using a method involving an unbiased model or method (Fenker and Zoota 2001).

2.1.2. Regression Method

Regression models can be used to assess a site's suitability for a retail location and have been used often by real estate professionals (Fenker and Zoota 2001; Birkin and others 1996). The regression method seeks to calculate a score for overall site attractiveness, the dependent variable, from independent variables which are believed to be positively and negatively correlated to sales potential for retail sites (Mitchell 2009). Independent variables used often include whether a property is on a corner, road visibility, the attractiveness to consumers of neighboring stores, competitor site proximity, road access, whether the retail location has a left turn signal, traffic counts, demographic profiles and more. Each independent variable in the regression equation is assigned weights of positive or negative coefficients reflecting the impact of each parameter on the site's likelihood for success or failure (i.e. sales potential). After applying the regression analysis, retail sites with higher scores are deemed to have the greatest likelihood of success.

The equations can be determined either mathematically using various regression techniques or intuitively by assigning variable weights subjectively (Fenker and Zoota 2001). Such models are applied with varied sophistication based on the mathematical and technological abilities of the particular buyer or real estate professional (Fenker and Zoota 2001).

2.1.3. Trade Area Method

Another method used to assess sales potential for retail sites analyzes the characteristics of the population within site's trade area. A trade area is the area surrounding a store's location from which patrons are likely to travel to the store; in other words, trade areas enclose the catchment population. Trade areas have been determined using many different methods.

Many of the trade area techniques use distance measures to characterize costs of travel to a store. As distance increases, so do the financial and intrinsic costs involved in patronizing a store location. Difficulty in getting to a store was historically approximated by a distance measure using a straight line radial distance drawn as a circular buffer (Birkin et al. 1996). Use of straight line radii for distance measures generally produced poor estimates of a trade area because this method fails to address the travel realities of road paths and natural obstacles, such as mountains or lakes, which require travel around or over these objects adding time and distance cost (Miller 2010).

As digital road data became more available, road network distance measures became preferred to straight line radial distances. Network distance measures more accurately reflect distance traveled to reach a location. While road networks better estimate distance they do not provide the best approximation of difficulty getting to a location (Birkin et al. 1996). Traveling a few miles may sound difficult to most people but anyone stuck in LA traffic who was running late for work with a few miles to go understands how frustrating and stressful traveling short distances might be.

Today, calculation of drive times using speed limit data in conjunction with road network data provide better approximations of the cost of reaching a retail site location. Time intervals are used to characterize the relative travel difficulty endured in order to reach a store. Typical examples of time intervals used in analysis are 3, 5, and 7 minutes for highly substitutable goods within urban and suburban areas. Willingness to travel is greater for specialty goods which are not readily available in most general or grocery stores so larger time intervals of 5, 10 and 15 minutes are typically used for analysis of these items. Trade areas created with drive time distance measures produce trade areas with greater distances from the site along freeways and major roads and shorter distances from a site along side streets and streets with slow speed limits (Birkin et al. 1996; Miller 2010).

Drawbacks of this method are that speed limits are not always abided by in light traffic and may not be achievable in heavy traffic. Additionally, traffic fluctuates greatly depending on the day of the week, time of day and with local school schedules. Further investigations into these variances are warranted but are outside the scope of this case study (Miller 2010).

For the purposes of this investigation, detailed drive times were found using Esri's Business Analyst to delineate trade areas for each store site. The next section provides an overview of the tools available in Business Analyst.

2.1.4. Esri's Trade Area Definition Tools

Esri's Business Analyst has many trade area definition tools which can be used to delineate a site's trade area. These seventeen trade area tools are summarized in Table 1. These tools allow users to define trade areas in a variety of ways, using methods ranging from simple to complex. Additionally, to help a user characterize trade areas, Esri's Business Analyst contains a substantial data resource about businesses, business performance, population, demographics, and consumer spending information, as well as Tapestry Segmentation data which provide consumer profile information. This software also allows for custom data layers to be imported. A custom population data layer was created and imported for this case study.

Table 1 Esri's Trade Area Tools and Descriptions

Tools	Description
Create Trade Area From Geography Levels	Generates trade areas based on standard geographic units.
Create Trade Area From Sub-geography Layer	Generates trade areas from the features of an input polygon layer that intersects a defined boundary layer.
Customer Derived Trade Areas	Creates trade areas around stores based on the number of customers or volume attribute of each customer.
Data Driven Rings	Creates a new feature class of ring trade area features. The radii are determined by a field in the ring center (store) layer.
Dissolve by Attribute Range	Aggregates and dissolves features based on specified attributes.
Drive Time	Creates a new feature class of trade areas, based on drive time or driving distance, around store point features.
Grids	Generates an equidistant vector based grid network for a specified area.
Huff's Equal Probability Trade Areas	Generates areas of competitive advantage boundaries between stores weighted on one or more variables. These weights can be calculated based on the results of a Huff Model.
Market Penetration	Calculates the market penetration based on customer data within an area.
Measure Cannibalization	Calculates the amount of overlap between two or more trade areas.
Monitor Trade Area Change	Creates a new feature class and report that analyze how trade areas have changed over time
Remove Trade Area Overlap	Removes overlap (cannibalization) between trade areas
Static Rings	Creates a new feature class of ring trade area features using a set of radii
Thiessen Polygons	Generates competitive advantage trade areas for each store by creating boundary lines equidistant from each of the store locations.
Threshold Data Driven Ring	Creates rings around stores. The radii of the rings are determined by expanding from the store location until they meet the criteria included in the store layer.
Threshold Trade Areas	Creates rings around your stores. The radii of the rings are determined by expanding from the store location until they meet your criteria.

(Miller 2010)

Esri’s trade area toolset utilizes geographic, distance and other data driven variables to delineate trade areas. Geographic features utilized to define trade areas include census tracts and administrative units. Distance measures are used to define Thiessen polygons, static rings, grids, and drive time trade areas. The rest of Esri's trade area tools incorporate distance with other supplementary variables such as store and competitor information, customer data, demographics and consumer spending information. Business Analyst is a robust tool that provides a user with all that is needed for most trade area definition tasks (Miller 2010).

2.2 Determining catchment population characteristics

Catchment population characteristics are calculated using existing census and other demographic and economic data which are widely available. The key challenge here is how to divide and redistribute the population counts and characteristics from the standard census and other polygons into the trade areas determined for a specific store location site analysis. In this section, the geography of census data is briefly summarized in order to explain the nature of aggregated census data. This is followed by a discussion of issues that arise from the use of aggregated data and some techniques that have been used to disaggregate it, including the use of land use data. Finally the concept of areal interpolation and specifically dasymetric mapping techniques similar to those used in this case study are introduced.

2.2.1 Census geography

Privacy concerns prevent census data from being released at the household level. Thus, household level information is aggregated to *blocks* which are geographic areas delineated in such a way that the total population is between 600 and 3,000 people and the demographic characteristics within the block are somewhat homogeneous. Blocks are themselves aggregated into larger *block groups* and block groups into *census tracts.* Importantly, for ease in some kinds of trade area analyses, blocks are also often reduced to centroid point features called *block points*

(Peters and MacDonald 2004). Block level data tables contain only population counts. Full census data are provided at the block group and larger aggregates.

2.2.2. Issues related to the use of aggregated data

Household count data can be aggregated into an infinite number of different size and shape polygons, each of which may be just as valid due to spatial autocorrelation. Spatial autocorrelation is the idea that sampled geographic data will likely be more similar to that from nearby locations than from more distant ones. As stated by Tobler in 1970, "everything is related to everything else but near things are more related than distant things" (Tobler 1970). Thus when dealing with Census data, population characteristics are likely to be similar to others nearby due to a human tendency to group near like individuals.

However, geographic boundaries can be manipulated to produce population characteristic distributions that are favorable or unfavorable to a specific end. Aggregation of geographic data results in the Modifiable Areal Unit Problem (MAUP) (Wong 2009). Congressional District gerrymandering and research conducted un-objectively to support desired outcomes are examples of such intentional manipulations related to the MAUP (Kelly 2012). Many dissimilar instances grouped together in arbitrary or manipulated ways also lead to research outcomes and real world outcomes that are biased and unrepresentative of realities.

2.2.3. Collapsing aggregated data to polygon centroids for spatial overlay

When using census aggregates above blocks (e.g. block groups or tracts) as the source layer for a spatial overlay, Business Analyst uses a method called weighted block centroid retrieval. Here each block centroid is assigned a proportion of its higher level aggregate's data values based on each block's population as a percentage of the enclosing census aggregate's population. Once population values are assigned to block points, during the overlay process, values from the source layer are included in the target layer results based on inclusion of block points within the target polygons. According to Business Analyst help documents on Spatial Overlay this will be more accurate than simple centroid inclusion retrieval method using the larger census aggregate centroids.

Using aggregated data collapsed to the centroid *block points* as done by Esri's weighted block centroid retrieval can still be problematic. In an overlay analysis of spatially incongruent layers, "centroid containment" rules for inclusion of source layers features in an overlay result is the most basic form of dealing with features having spatially mismatched boundaries (Miller 2010). For example, if a block centroid falls outside of a trade area, the entire block population will be treated as not intersecting the trade area even though in reality some of the block population may fall within the trade area. Conversely, a block polygon that intersects only a small sliver of the trade area but whose centroid is within the trade area's extent produces outputs reflecting complete inclusion of all block population from the source layer in the result. When applied to population polygons, entire populations are accounted for in the features intersecting the centroid and no population is reflected for features where the intersecting layer does not intersect the centroid despite potentially intersecting a majority of the population aggregate polygon (Miller 2010).

Using centroid containment for inclusion and exclusion in analysis is undesirable unless the target areas of a spatial overlay are large relative to the census aggregates. Large errors will occur when source and target features are similar in size (Ignizio and Zandbergen 2010). Methods which are more advanced than rules for inclusion and exclusion in addressing spatially different polygons in analysis are commonly referred to as areal interpolation according to Goodchild and Lam as cited in (Zandbergen 2011).

2.2.4. Methods of areal interpolation

The most basic form of areal interpolation is areal weighting which weights population included and excluded in a source zone by the ratio of its intersection to the source layer feature area. Another form of aerial interpolation is “surface fitting” where a surface is fitted to the data in source areas and typically inferential statistics are used to interpolate values (Zandbergen 2011).

Another method of areal interpolation is dasymetric mapping in which ancillary data are used to distribute population unequally within the source layer features. Dasymetic mapping is the process of disaggregating data into finer units of analysis to help refine locations of population or other phenomena (Mennis 2003). The results preserve known population within each source area in the target area results. This is referred to as the pycnophylactic property (Qiu and Cromley 2013).

2.2.5 Land use as auxiliary data to disaggregate population within census zones

Land cover is the ancillary data most often used to refine population distribution using dasymetric mapping. Dasymetric mapping using land cover data usually employs an overlay of population polygons with land cover classified data. Population is apportioned to varying land cover areas by assigning weights to their land cover classifications. Importantly, some land cover classifications such as water or natural areas are weighted zero and not apportioned any population. The remaining populated areas are apportioned by aerial weighting. These aerial weighted polygons assume population distribution is uniform in target zones but since these areas are usually much smaller than the aggregates, the results are a more accurate estimate (Qiu and Cromley 2013; Zandbergen 2011).

Early on, spectral signatures in Landsat images were used to spatially delineate various kinds of land cover (Amaral et al. 2012). Training areas with known land cover were used to determine the recorded spectral signatures for these known land cover types. These spectral signatures were then used to classify land cover throughout the Landsat images.

Another approach is to use remotely sensed thermal images to classify land use and land cover by comparing heat emissions from morning and night images (Wen and Yang Xiaofang 2011). Water bodies produce almost no heat emission so they are easiest to identify. Increasing levels of daytime heat emission are recorded from undeveloped land, residential and commercial areas. Industrial areas and high rises produce the largest heat emissions during the day. Residential areas tend to have a much greater contrast of heat emission levels between day to night as people are home and use energy more frequently at night. Commercial and industrial areas display the opposite effect as these areas are not typically operational at night.

Zanbergan and Ignazio (2010) used census block group population with large scale land cover data as the ancillary data to estimate population. Actual census block group population counts were then compared against calculated block group population to see the error produced. The authors used areal weighting, land cover, total imperviousness, imperviousness above 75 and 60 percent, "cleaned imperviousness" total roads, local roads and nighttime lights datasets to conduct dasymetric mapping. Imperviousness refers to the imperviousness of surfaces to show where population is likely with population more likely on the least impervious surfaces due to paving and structures likely in these areas. Similarly road density is another surrogate for the presence of population.

Land cover and imperviousness performed the best among the datasets used with the lowest errors produced. Zandbergen and Ignazio found errors ranging from 11.9 to 14.5 using landcover data. Later in 2011, Zanbergen augmented his previous study and added address point and residential address point data to the same analysis. Again he found land cover had a similar error, this time at 11.6 percent, while imperviousness performed better at 10.8 percent.

Far outperforming the other datasets, address points had only 4.9 percent error and residential address points produced only 4.2 percent. For this reason, a dataset of only populated developed residential parcels called “Populated parcels” in short was created and used in this study. To produce this dataset census aggregated populations were areally weighted to residential areas with existing structures.

2.3 Background Summary

Methods for estimating a store’s sales potential by analogue, regression and trade area methods were illustrated. Esri’s trade area creation tools were described. Census geographic data was overviewed and issues arising from using aggregated data were discussed. Methods of areal interpolation used in past research to disaggregate aggregated data including dasymetric mapping were reviewed.

Calculation of trade area catchment population and demographics used to estimate store sales potential depends on trade area definition and population most often provided as census aggregates. A methodology of dasymetric mapping to disaggregate population as well as methods of trade area creation used by this study to find differences in calculated catchment population and demographics are reviewed in the following Chapter.

CHAPTER 3: DATA AND METHODS

Analyzing a retail store's trade area population is crucial to determining the potential of a retail site with the results affecting decision making for constructing, locating or relocating, or maintaining a retail site. A case study was conducted using different aggregation levels of population data and various methods to create store site trade areas to show the effects on calculated trade area population counts and related population attributes. The accuracy of results will have an effect real estate decision making based on the calculated characteristics of catchment population.

Potential retail sites located in various parts of San Diego County with different population densities were selected for this case study to see if results varied by population density. This case study explores how trade area population characteristics found by using different population estimates from aggregated Census population data and concentrated population areas impact the calculation of catchment population and related characteristics. A generalized workflow for this case study is shown in Figure 1.

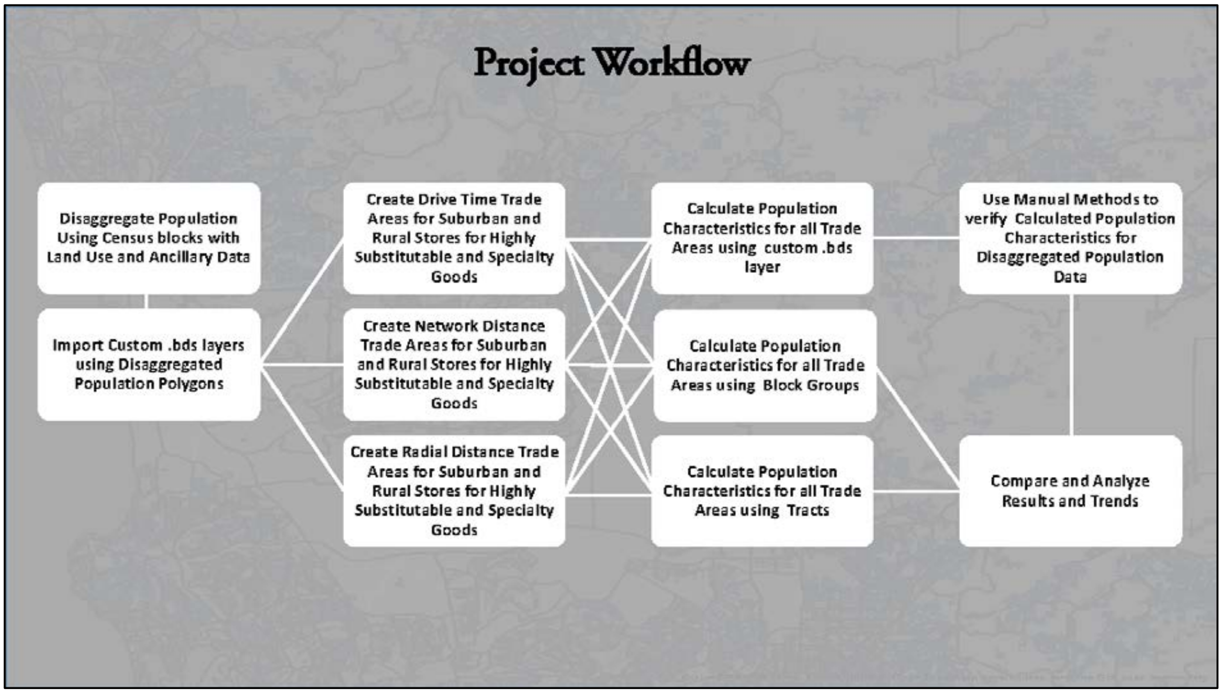

Figure 1. Project Workflow

3.1 Geographic Data Sources

Data for these analyses were collected from Esri Business' Analyst, the US Census, and SanGIS/SANDAG regional data warehouse. These data sources provide data in the form of tables, shapefiles and geodatabase for boundary and attribute data. When data layers were used in this spatial analysis, Esri's Business Analyst automatically converted data layers from differing datum and projects them to be compatible with the base layer datum.

3.1.1 Assessor Parcel Data

The Assessor Parcel Data produced by SanGIS contains boundary files for each tax parcel in San Diego County. Each parcel has a unique assessor parcel number (APN) and the corresponding parcel's legal boundary and ownership information. Each APN number is unique to ownership entity or group, however boundaries of APNs in the case of condominiums, which all occupy the same parcel boundary, are overlaid resulting in multiple ownership entities or groups for a single geographic boundary. Alternatively multifamily housing (Apartments and Senior Living Facilities) while having many residents generally have one owner for one geographic boundary. In the case of both condominiums and apartments, one parcel boundary can have many households living in a single parcel boundary.

Single family residences have a one to one parcel boundary to household ratio. This results in different parcel areas having one household and one owner, many households and one owner or many households and many owners. “Populated parcels” were areally weighted by the percentage of total area for “Populated parcels” within each block. The corresponding population for each block was multiplied by these weights to arrive at a new estimate of population for each polygon with source layer population preserved in target features referred to as the pycnophylactic property of dasymetric mapping. This new layer assumes that population distribution is uniform across these smaller polygons. However, this resulted in large single family parcels of wealthy households being allocated the same population of a multifamily parcel the same size, especially where estates are large and a large area of land surrounds the residence. This means that a large estate may have received a population estimate similar to an entire apartment complex which has a much greater population despite a similarly sized area thus skewing population estimates.

3.1.2. Retail Analysis Data Demands

Retail site data utilized by most retail developers and analysts needs to be obtained easily and at a low cost. Esri’s Business Analyst provides generalized albeit rich data about consumers. Generally, survey data provide more detailed information about a specific retail site, such data are more time consuming to obtain and beyond the needs of most retail developers. Based on my experience in my office with current level of data used in real estate, data provided by Esri’s

Business Analyst are sufficient for most analysis needs and for this reason utilized in this study. Retail site data for this case study are provided by Esri's Business Analyst.

3.1.3. Land Use Data

Land Use data were provided by SanGIS's regional data warehouse. These data show the land use classification for each parcel in San Diego although contiguous APNs with the same land use classification are aggregated into larger land use polygons. There are several different designations within each classification of residential, commercial, retail, industrial, open space, conservation and more. For this analysis only the residential areas of single family, multifamily, condominium, student housing, were used to better classify population distribution. Other residential distinctions such as prisons, hotels, and others were omitted.

3.1.4 Population and other Demographic Data

Population and corresponding demographic data for this case study were taken from both Esri's Business Analyst and the US Census. Population data obtained from Esri's Business Analyst are from the biennial census as well as the American Community Survey. Non-census year data are derived from Esri's own models which project population totals and demographic characteristics. Esri's Business Analyst has data from the Census for Tract and Block Group Census boundaries. Block level data from Esri's Business Analyst are available but block boundaries are not given and blocks are represented by block centroids called block points. In order to account for the boundaries of Census Blocks, data obtained from the Census website directly provided the Block Boundaries and population counts within those blocks.

3.1.5 Developable Land Data

The county of San Diego develops many layers of ancillary data for use by its employees and the public. One of the layers produced by San Diego County is the “Developable Land” layer. This layer is comprised of areas that the county deems developable based on favorable topography not prohibitive to development costs and that is currently vacant. Land which may be undeveloped but having topography that is not relatively flat or with zoning restrictions or conservation protections are not considered favorable and are excluded from this layer.

3.2 Case Study Store Sites

The four store sites in San Diego County chosen for this study are pictured in Figure 2. Two of these store sites are located in suburban city areas (San Diego and Poway) and the other two are in cities in more rural areas of the county (Alpine and Ramona). As described below, each of these cities is unique in population, demographics and bordering communities. They were chosen to provide a wide range of conditions over which to test this methodology.

Figure 2: Store locations in suburban and rural areas of San Diego County. Suburban store locations are in the city of San Diego and Poway, rural locations are in the cities of Ramona and Alpine. Study area for data extent is outlined in blue.

The San Diego store site is located in the affluent community of Carmel Valley which is mostly comprised of single family residences and condominiums. The even more affluent cities of Del Mar and Rancho Santa Fe with multimillion dollar beach homes and equestrian ranches lie to the west and north respectively. To the east is the city of San Diego proper with many new single family home developments which were constructed prior to the housing crash with graded lots and open space awaiting future development. To the immediate south is a large commercial area home to Qualcomm’s ever growing campus and other commercial buildings with few apartment communities throughout the area. Even further to the south are the city of La Jolla and the Miramar marine base. The San Diego store location is near the coast to the west and has convenient access to Interstate 5 which connects to Interstate 805 to the south as well as Hwy 56 to the east.

The second suburban site located in the city of Poway is similar to the San Diego site in density but is slightly less affluent due to being more inland where residential real estate costs are slightly more affordable. The city of Poway and the Poway store are located next to Interstate 15 and Hwy 56 allowing for quick travel to the north, south, and west and Poway Road a major arterial road allows for quick travel east.

The third store located in the city of Alpine is considered a more rural location. Alpine is located close to Interstate 8 allowing for quick travel to the east and west, however Alpine is the easternmost city in San Diego County and surrounded by Native American reservations and hilly topography with sparse population to the north east, and south. Past the Native American reservations, a few miles to the west are the outskirts of the San Diego County suburban area which are more affluent and densely populated relative to Alpine.

Lastly, the fourth store is located in the rural city of Ramona. Ramona is located in the hilly region in the north central part of San Diego County just west of the city of Escondido, which is the north easternmost city on the edge of San Diego County's urban sprawl. Ramona has one access road leading into Escondido. However that road is a rural road and is time consuming to travel. As a result, Ramona is a city largely isolated from the rest of the county and getting in and out of Ramona is difficult. Ramona is surrounded by sparsely populated hilly areas in all other directions further isolating the city.

3.3 Determining Catchment Population

Catchment populations, the population within a store's trade area, are often modeled by looking at where people have their residences, but it might also be derived based on where people work or by taking into account travelers and commuters. While different ways of defining catchment population distribution are important and deserve further exploration, for the purposes of this case study the catchment population is determined using three different datasets. The first two are the traditional approach using the Census level aggregations, in this case census block groups and tracts. The third approach uses parcel and land use data to concentrate the census population aggregates into truer representations of where population most likely resides by removing those areas within census aggregates where population does not live. The next section describes the disaggregation process used here.

3.3.1 Disaggregation of Census Population Aggregates

Disaggregation of census population aggregates for the third approach to determining catchment population was achieved through the use of Land Use and other ancillary data from the county. This required a multistep extraction process.

First, using the county's "Land Use" layer, polygons with residential land use designations of multifamily, condominium, single family residences, dormitories and military housing polygons were isolated. Some other residential land use classifications such as hotels, hospitals were not included as residential land use, despite the county classification of residential land use, as they are not counted toward population by the Census. Prisons were also omitted from inclusion despite their residential classification by the county due to prison population lacking the mobility necessary to patronize a retail store. While prison areas were removed, their population is included in Census data. These populations were not removed from the analysis here, but since they are small numbers relative to the general population, inclusion of these numbers in the total figures used is not considered an issue of concern. Also, since most San Diego prisons are in the southern portion of the County, south of the study area, there should be minimal impact on results. The isolated residential areas were extracted into a new feature layer called “Residential land use”.

As noted earlier, land use polygons are aggregations of contiguous assessor parcels of the same land use. Assessor parcel boundaries are completely within or identical to land use boundaries. As a result, it is possible to isolate each assessor parcel within each land use polygon. Thus, the next step involved overlaying “Residential land use” polygons on the assessor parcels. The intersecting features were extracted and exported to a new feature layer called “Residential parcels.”

Residential land use means that residential uses are allowed but this does not indicate whether an area was developed or undeveloped. Undeveloped areas do not have residences despite the residential land use designation. Thus, to ensure that the parcels to be included in this analysis have structures on them and were not merely designated residential, undeveloped parcels were identified. The “Developable Land” layer created by the county containing parcels with potential for development, but not yet developed, was utilized. “Developable Land” was overlaid on “Residential parcels” and the intersecting areas were removed as these areas were currently undeveloped. The remaining parcels was exported to create “Developed residential parcels” layer.

“Developed residential parcels” does not yet completely indicate the existence of a residential structure on a parcel as some vacant areas within residential land use polygons are not considered developable by the county due to issues of fitness for development. Land fit for development is relatively flat, currently undeveloped, with permissible zoning and not subject to conservation protections. Parcels which do not meet these requirements despite falling within residential land use zoned areas cannot be extracted using the “Developable Land” data. Thus, Census block data were used to further remove such unpopulated areas.

Census data collected about individual households are aggregated initially to census blocks and combined to form higher levels of aggregation. Census blocks are contiguous for all United States geographies including uninhabitable areas such as water bodies, conservation areas, mountainous or hilly areas or valleys where residences are nonexistent or very sparse. As a result, it is possible to use unpopulated census blocks to further identify unpopulated parcels.

This was accomplished by isolating census block polygons which have no population. Census polygons do not share exact boundaries with the assessor parcels. In order to create shared boundaries for both layers “Developed residential parcels” were split along all Census block boundaries. New polygon boundaries were formed when “Developed residential parcels” were located in more than one Census block. The resulting dataset assigned each new “Developed residential parcels” area with its corresponding block attribute. All “Developed residential parcels” area which intersected census blocks having no existing population were removed. The remaining polygons, referred to as “Populated parcels”, have both population and existing residences. Figure 3 shows an example of the result of this process.

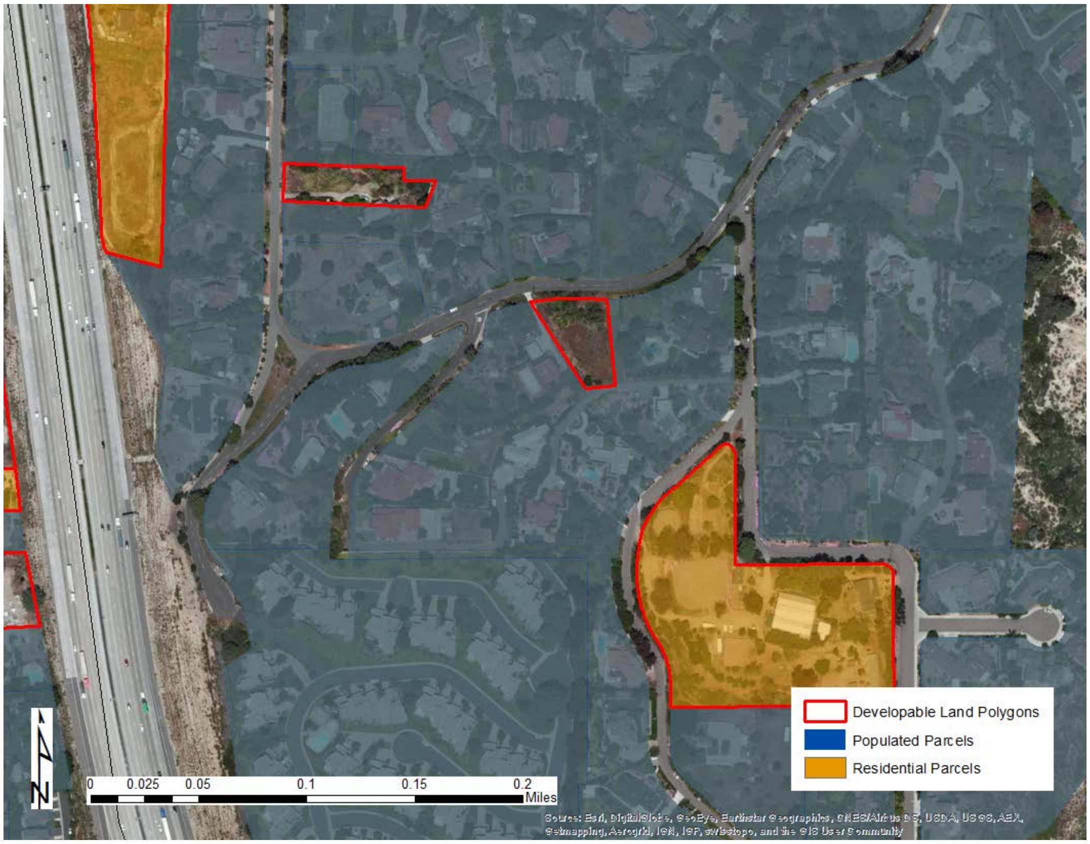

Figure 3: “Developable Land” (outlined in red) overlain on “Residential parcels”. Here the final “Populated parcels,” shown as blue areas, overlay the more extensive yellow “Residential parcels” which shows through the clear developable land polygons. The Developable Land polygons in this image are examples of polygons removed from the “Residential parcels” layer.

These remaining polygons provide a more accurate consolidated representation of where census populations reside than the census aggregates which contain many areas where people do not live. It was then necessary to apportion population to the “Populated parcels” from the block aggregates. In order to do this, the percent of each block polygon area occupied by each of the “Populated parcels” within each census block was calculated. Population for each resulting parcel polygon was then apportioned by multiplying the block population and household count by the calculated area percentage. All area percentages for each “Populated parcels” within a block when added together by block summed to one.   The result is population for each block aggregate allocated by area to the existing “Populated parcels” within each block, providing a more concentrated representation of each census block’s population distribution.

Weighting population and households by area of these polygons is only one method of disaggregating population and has many shortcomings. For example, some densely populated areas like apartment complexes or high rise areas which have much more population are allocated here the same population as rural areas that contain only one house for an area similar in size to a nearby apartment community. Accounting for these differences is worth further investigation but not in the scope of this study.

3.4 Calculation of Detailed Drive Time Trade Areas

Drive Time Trade Areas were constructed in Business Analyst using the Business Analyst Trade Area tools. For the purposes of this study, Drive Time Trade Areas were estimated around store locations with varying densities of population selected in San Diego County. Drive Time Trade Areas were created using Drive Time Trade Areas of 3, 5, and 7 minutes per site. The Detailed Drive Times option was selected so that only the road network able to be traversed within the selected time frames was included, excluding areas not reachable. This results in a more precise output. If detailed drive time areas are not selected, the areas covered by road networks are joined by distant end points to form larger polygons which may include areas which may not be reachable within the allotted time.

3.5 Calculation of Radial and Network Distance Trade Areas

Trade areas were constructed in Business Analyst using the Business Analyst Trade Area tools for Simple Rings and Drive Time distance. For the purposes of this study, distance Trade Areas were estimated around all stores described in Section 3.2. Trade Areas were created using radial and network distances of 3, 5, and 7 miles for each site.

3.6 Calculation of Trade Area Population and Characteristics

Esri's Business Analyst Drive time tool allows for the creation of reports when a drive time area is calculated summarizing the drive time area with population and consumer spending characteristics. In lieu of this method, for this project spatial overlay was used to find the catchment population and corresponding characteristics for each drive time trade area. Spatial overlays were also used to calculate the population and characteristics of the radial and network distance Trade Areas defined in the previous subsection.

When conducting these overlay analyses, Esri's Business Analyst automatically used two different methods based on the population dataset used as the source layer. When using 'Populated parcels', the overlay process apportioned the percentage of population equal to the area of the polygon within the trade area. When using the census aggregates as the source layer, Business Analyst used the weighted block centroid retrieval method described earlier. The results of these spatial overlay functions using the three different population data sources for all trade areas are discussed in the following chapter.

CHAPTER 4: RESULTS

As described in the previous chapter, for the study area used in this case study, shown in Figure 2 in the previous chapter, population data were concentrated to residential areas using a combination of land use parcels, supplemental county data, and census block data. The results of these operations are illustrated in Figure 4 which shows in blue the areas where “Developed residential parcels” were found.

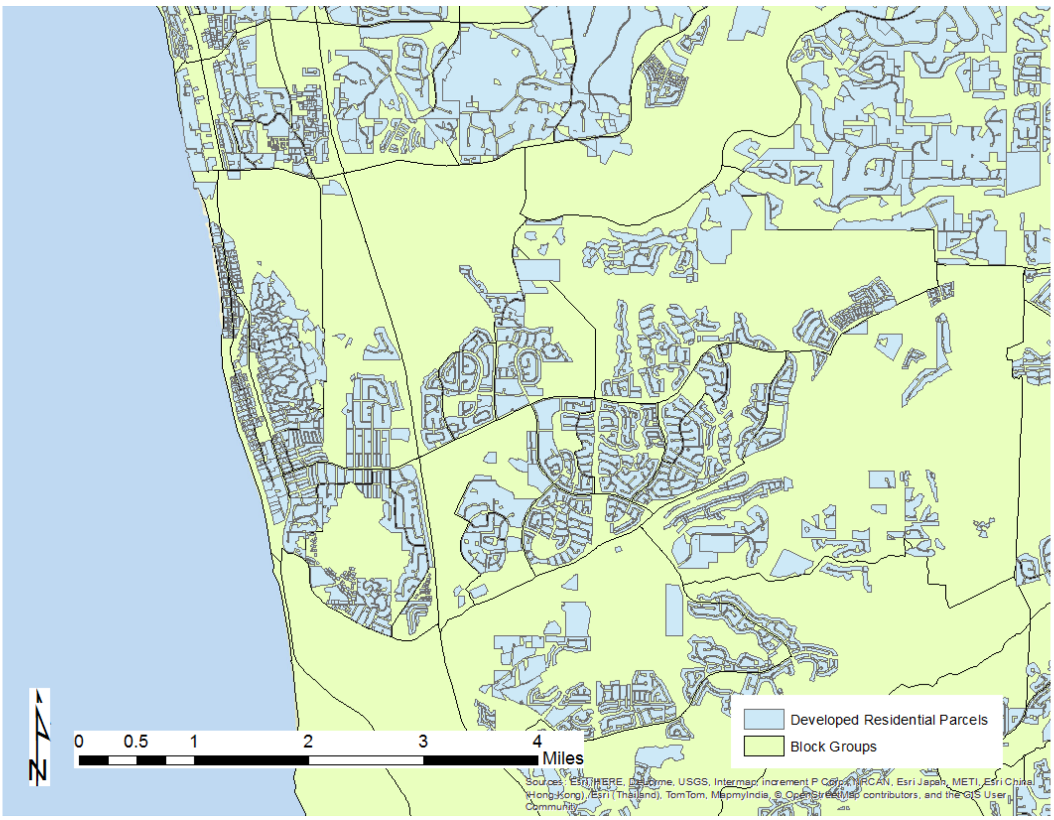

Figure 4 : “Developed residential parcels” in San Diego County. Supplemental data was used to identify developable land for all land uses. Developable land was extracted and removed from the Residential land use polygons. Remaining “Developed residential parcels” have residential land use designation and existing development.

Lastly Census blocks without population were also removed. The resulting polygons, illustrated in transparent blue in Figure 5, are referred to as “Populated parcels” for the remainder of this document. The blow-up in Figure 5 with “Populated parcels” showing in a transparent hue allows visual verification of the correct classification of these areas when overlain on satellite imagery. The areas in Figure 5 in yellow are uninhabited areas within census blocks with population. Figure 6 shows just the “Populated parcels”, in blue overlaid on an aerial image to illustrate more clearly where these are located.

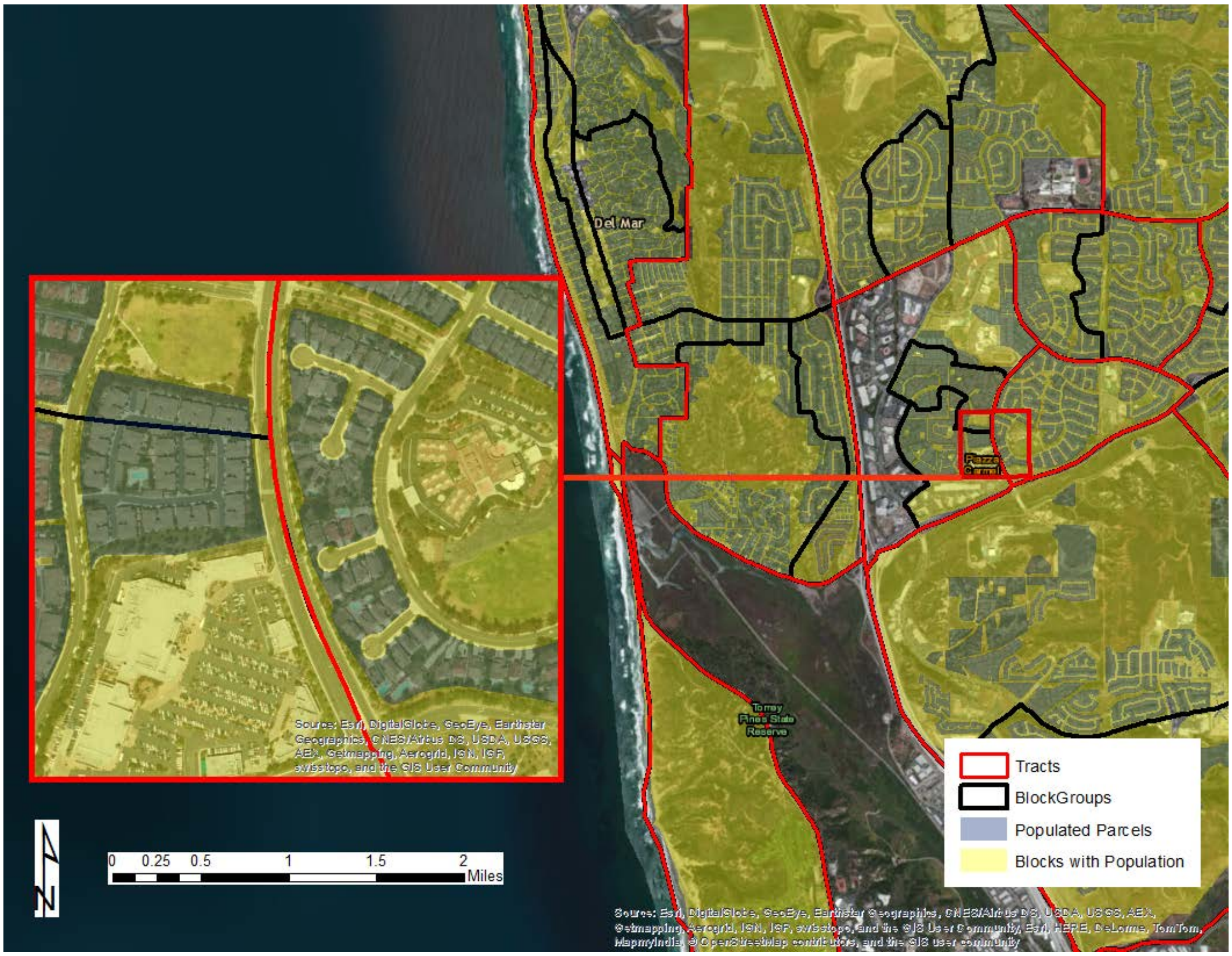

Figure 5 : Polygons with existing residential land use development aka “Populated parcels” overlaid on yellowish background areas representing all Census Blocks. Census Block Group boundaries are shown in Black and these are overlain by Census tract boundaries in red. The extent of the yellow area shows that a much larger area is included in the Census zones whether or not there is any population contained.

Figure 6 : “Populated parcels” showing imagery of underlying residences.

4.1 Results of Drive Time Trade Areas

For the first portion of this case study trade areas were formed around store sites using detailed drive times for 2 different product type scenarios. The first scenario was for a store with non-specialty goods and services meaning consumers would be willing to drive smaller distances to obtain these items than more specialized items. For this scenario drive time areas were created for each store at three, five, and seven minute intervals. Figures 7, 8, 9 and 10 show the corresponding drive time trade areas for the four store sites in San Diego County.

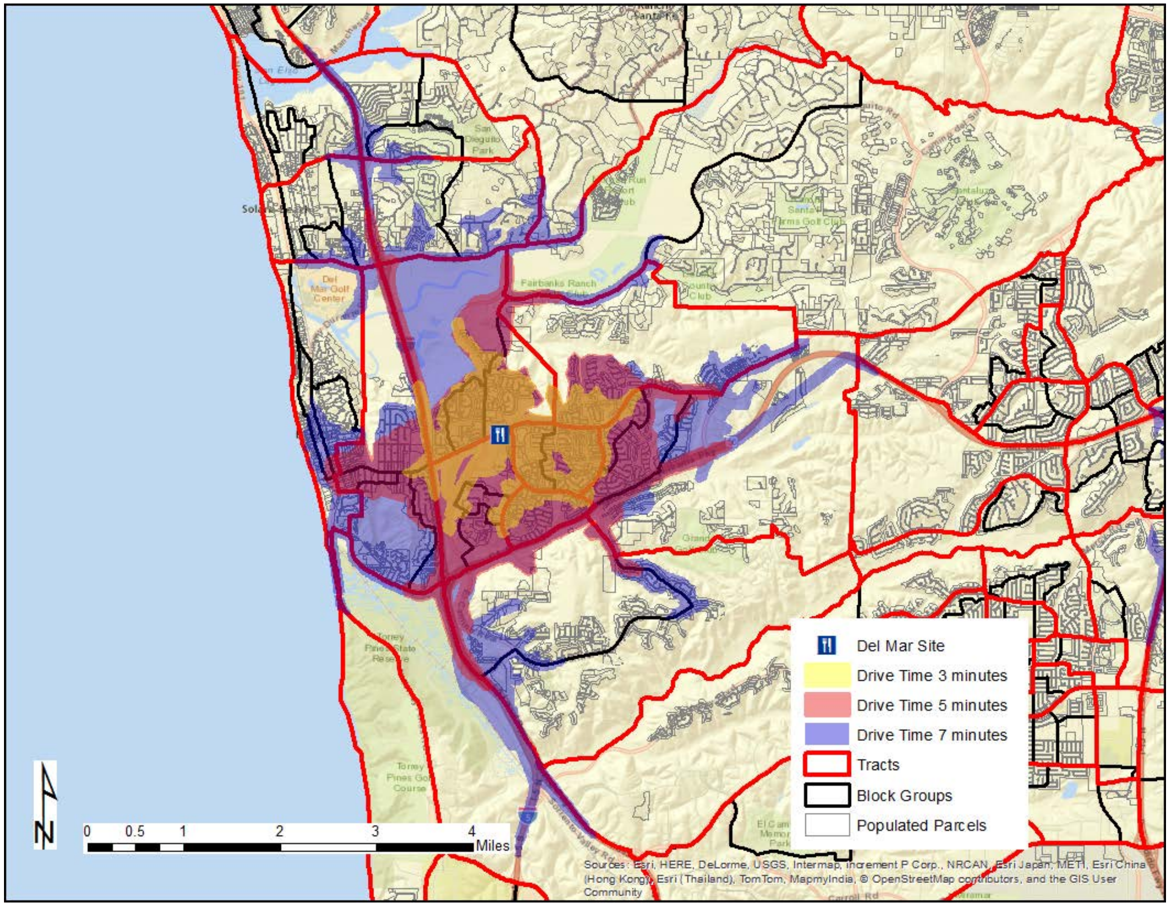

Figure 7 : Trade Areas using detailed drive times for three, five, and seven minute intervals.
San Diego (Del Mar Heights Rd.) store

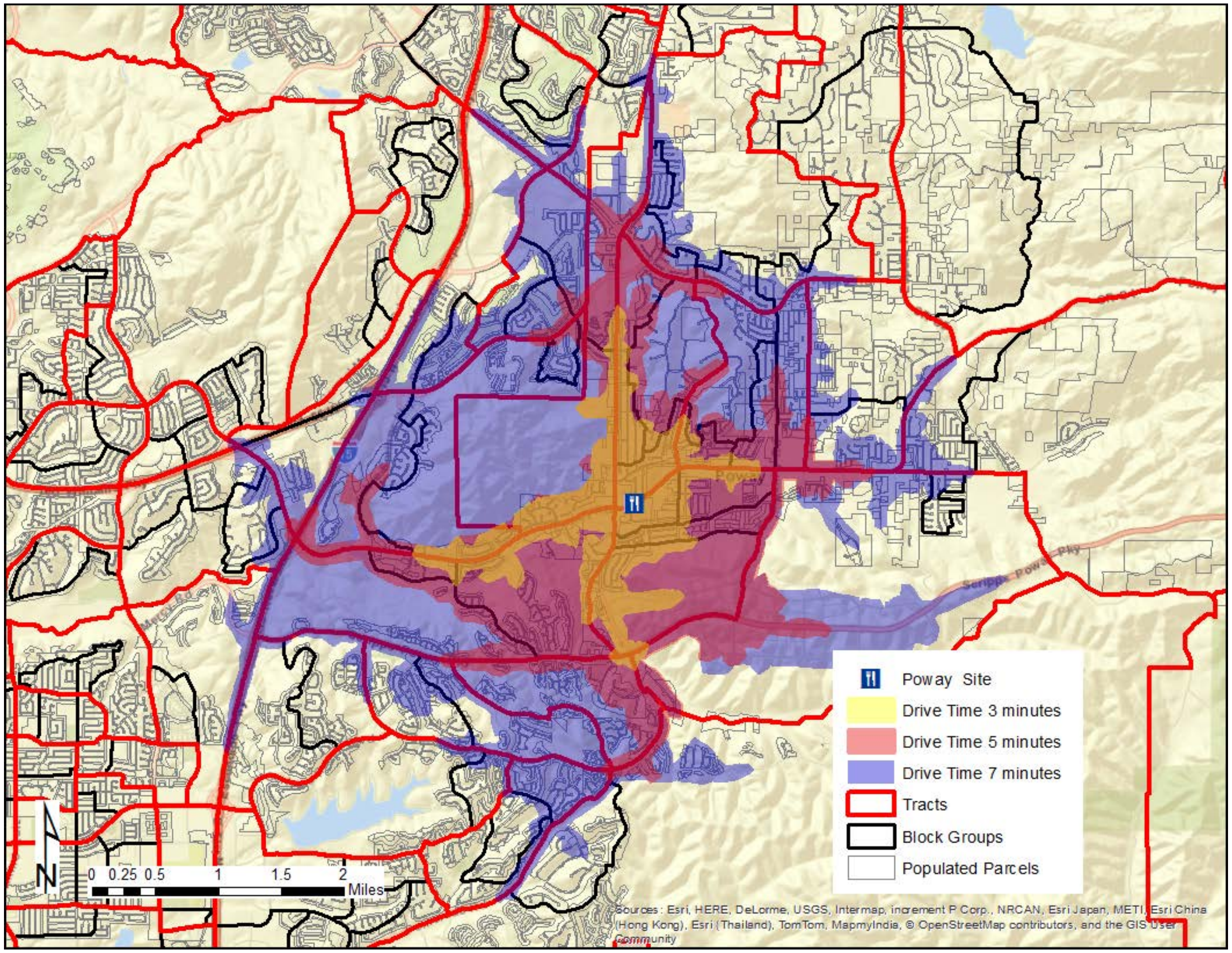

Figure 8: Trade Areas using detailed drive times for three, five, and seven minute intervals. Poway store;

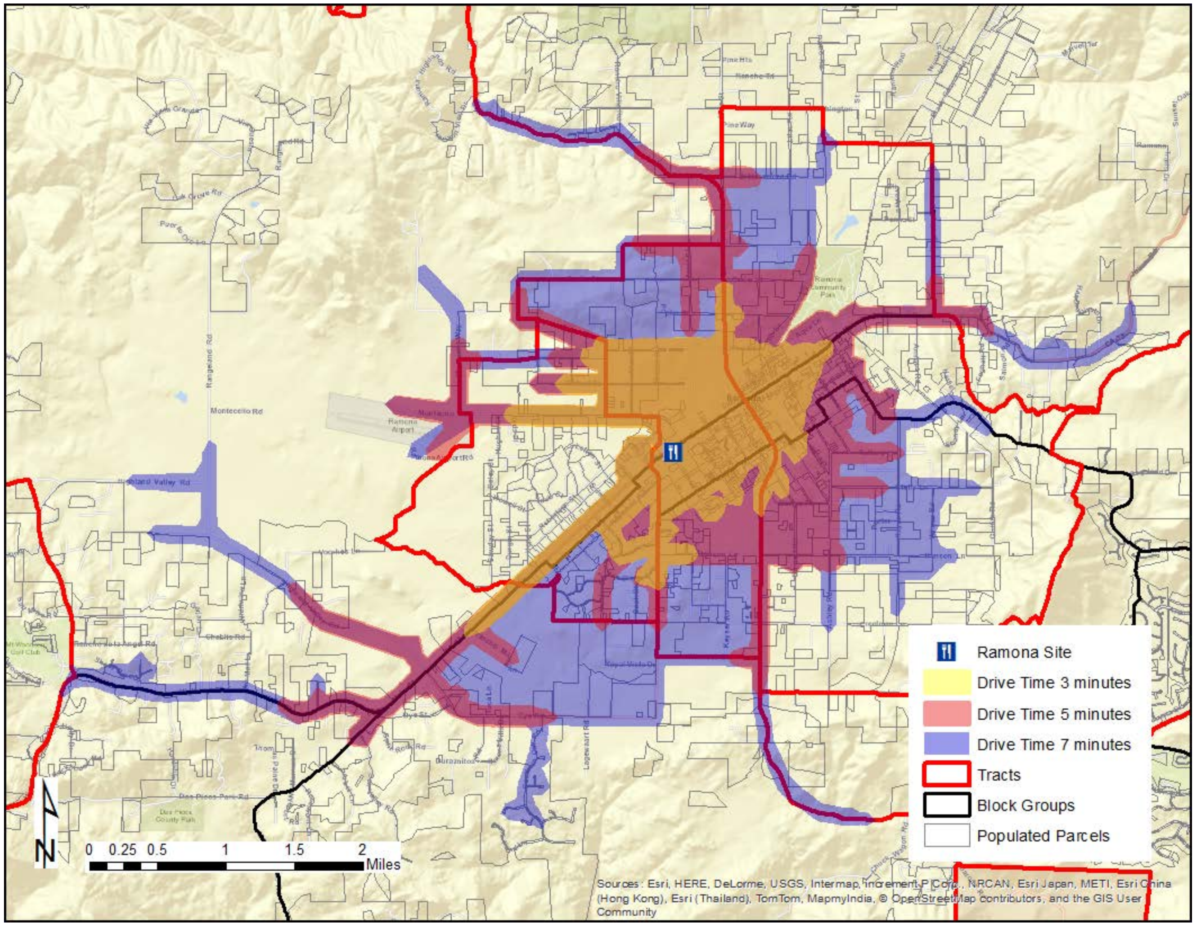

Figure 9 Trade Areas using detailed drive times for three, five, and seven minute intervals. Ramona store

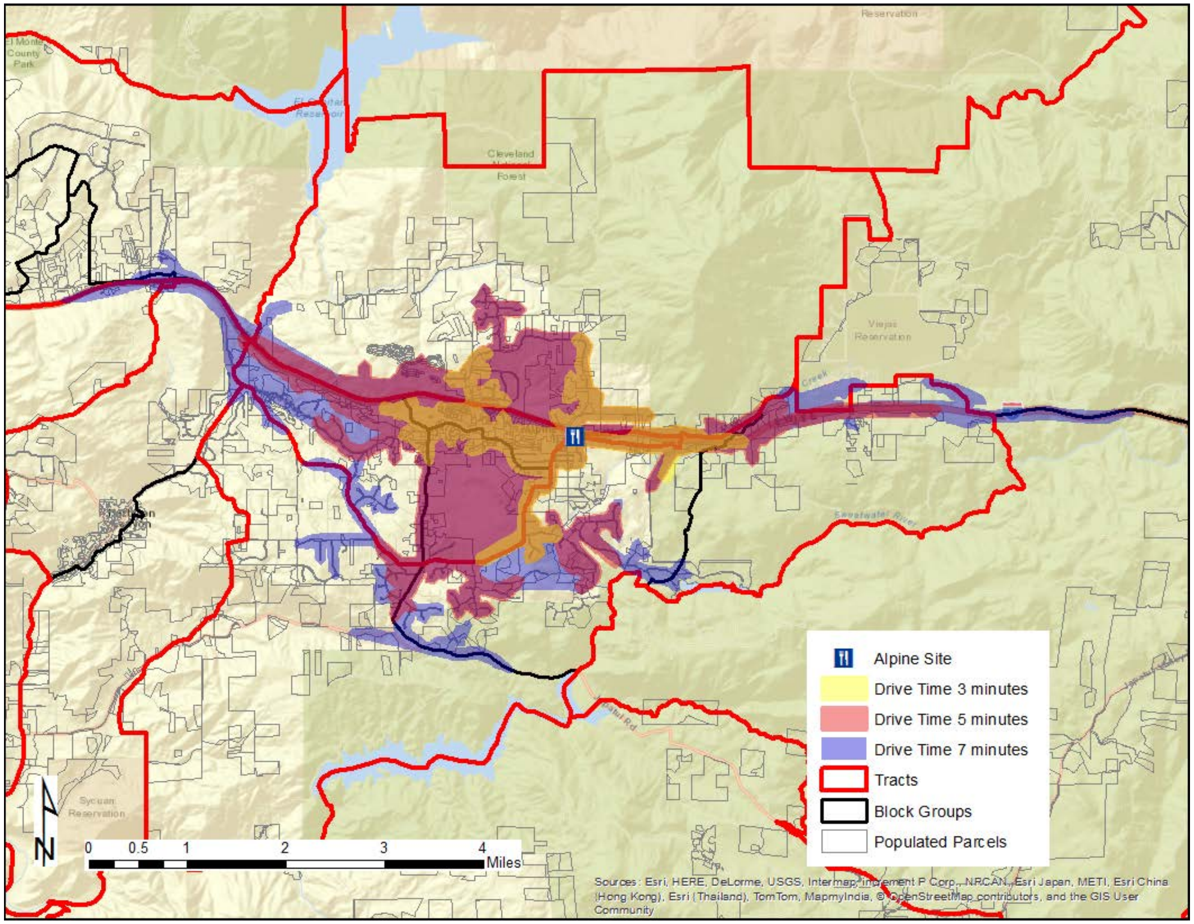

Figure 10: Trade Areas using detailed drive times for three, five, and seven minute intervals. Alpine store

To account for stores that sell more specialty goods which customers would be more inclined to endure greater travel difficulty to procure, additional drive time areas at time intervals of five, ten, and fifteen minutes were created. These larger trade areas, shown in Figure 11 were created for the westernmost location of San Diego County and two easternmost locations eliminating the Poway location. If the Poway location were included drive time areas for the suburban locations which overlapped would cause errors in the spatial overlay results for this study. For this reason the Poway location was omitted from analysis of these larger drive time distance thresholds.

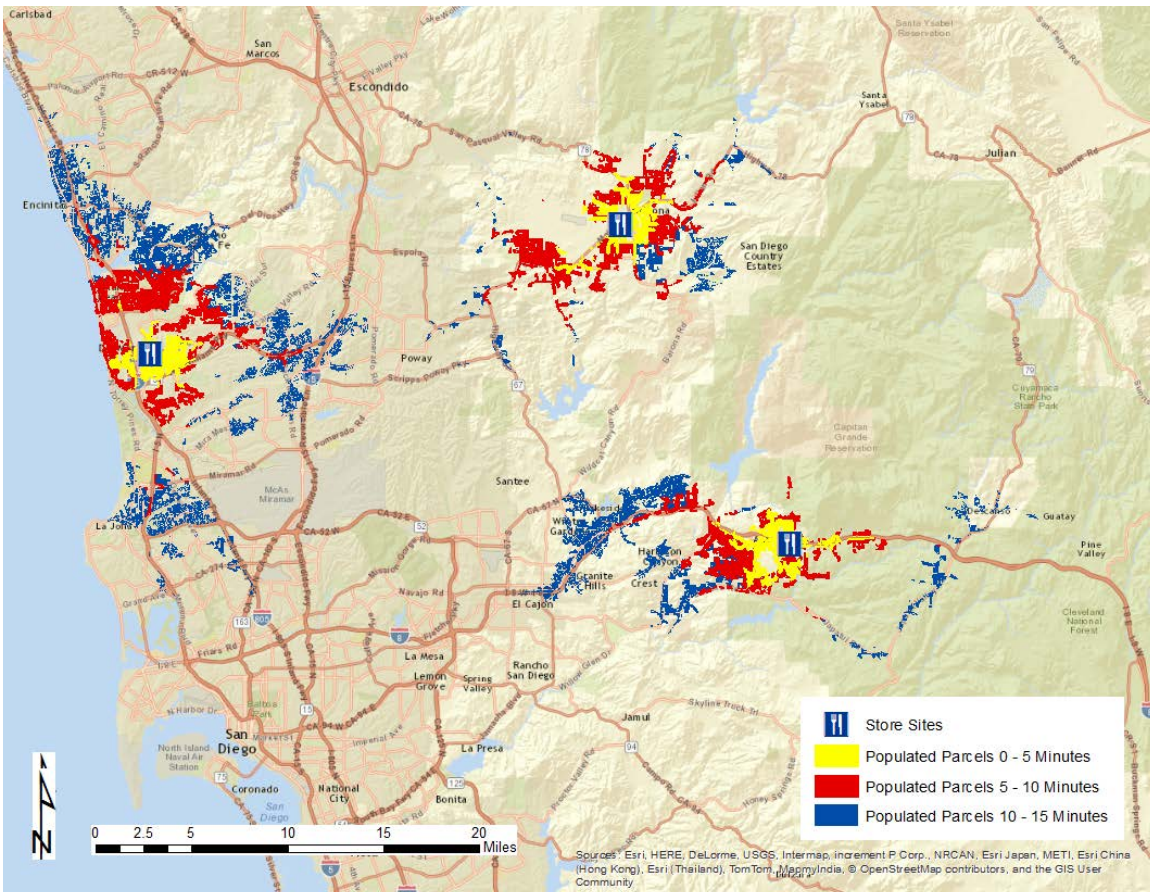

Figure 11 "Populated parcels" intersecting detailed drive time areas of 5, 10, and 15 minutes for the San Diego, Ramona, and Alpine store sites. The suburban San Diego location has noticeably more residential parcels than the more rural Ramona and Alpine.

This case study explored the effect on results of calculations of catchment population and demographics using concentrated population data as opposed to census aggregates. The effects on the results could impact retail real estate decision making. Spatial overlay was used to find the catchment population characteristics for the Drive Time Trade Areas using the “Populated parcels”, Census Block Groups, and Tracts. Differences in the calculated catchment populations using the 3 base population layers were analyzed. In order to conduct the Spatial Overlay a custom .bds (Business Analyst Dataset) layer was created using the “Populated parcels” described in Chapter 3 along with the layer’s derived census attributes for each population polygon. The “Populated parcels” were imported and the population and households for each polygon were apportioned by area. All other attributes were joined to the associated block group level and weighted by either population or household depending on the normalizing metric. The normalizing metric is a count such as population or households for which measures such as per capita income and average household income are derived for Census data (Peters and MacDonald 2004). Block Group and Tract data from Esri’s default pre-packaged .bds layers were used for Spatial Overlays at the corresponding aggregate population level.

All detailed drive time trade areas for each store were calculated individually for each store and time increment in order to ensure that Spatial Overlays would be calculated as an aggregate from store location to the outer extent of the detailed drive time trade area. This means that despite overlapping of trade areas of different distances all population within the trade area boundaries was used for the calculations for each trade area. The results of the Spatial Overlays are shown in Table 2 for the grouping of three, five, and seven minute trade areas. The “Populated parcels” overlaid in the Spatial Overlay process are pictured in Figure 6 through Figure 9.

Figures 12 through 15 show all “Populated parcels” intersecting detailed drive time trade areas of three, five and seven minutes for the store locations. These land use polygons are overlaid on a less vibrant drive time trade areas of three, five and seven minutes of the same color. Vibrantly colored areas indicate areas within the store’s trade area where population lives. “Populated parcels” not within the trade area for each site are shown with a gray outline.

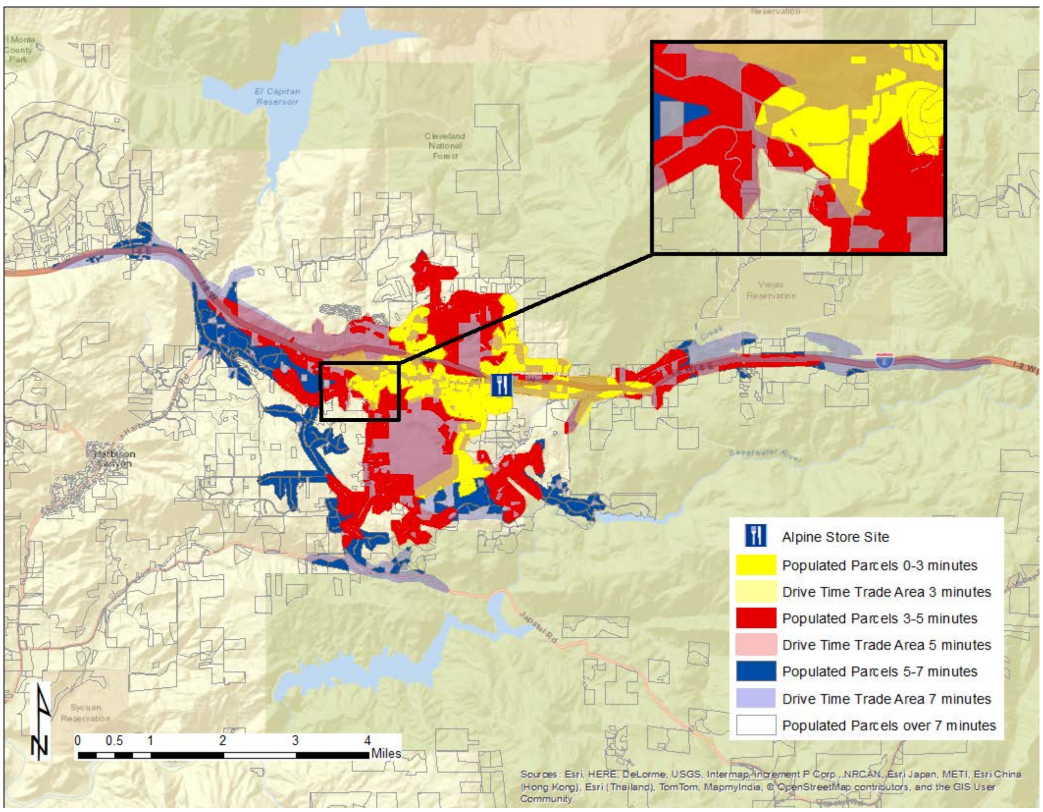

Figure 12 : All “Populated parcels” intersecting detailed drive time trade areas of three, five and seven minutes for the Alpine location. These “Populated parcels” are overlaid on a less vibrant detailed drive time trade areas of three, five and seven minutes of the same color. Vibrantly colored areas indicate residences. “Populated parcels” outside the trade area for the site are shown with a gray outline.

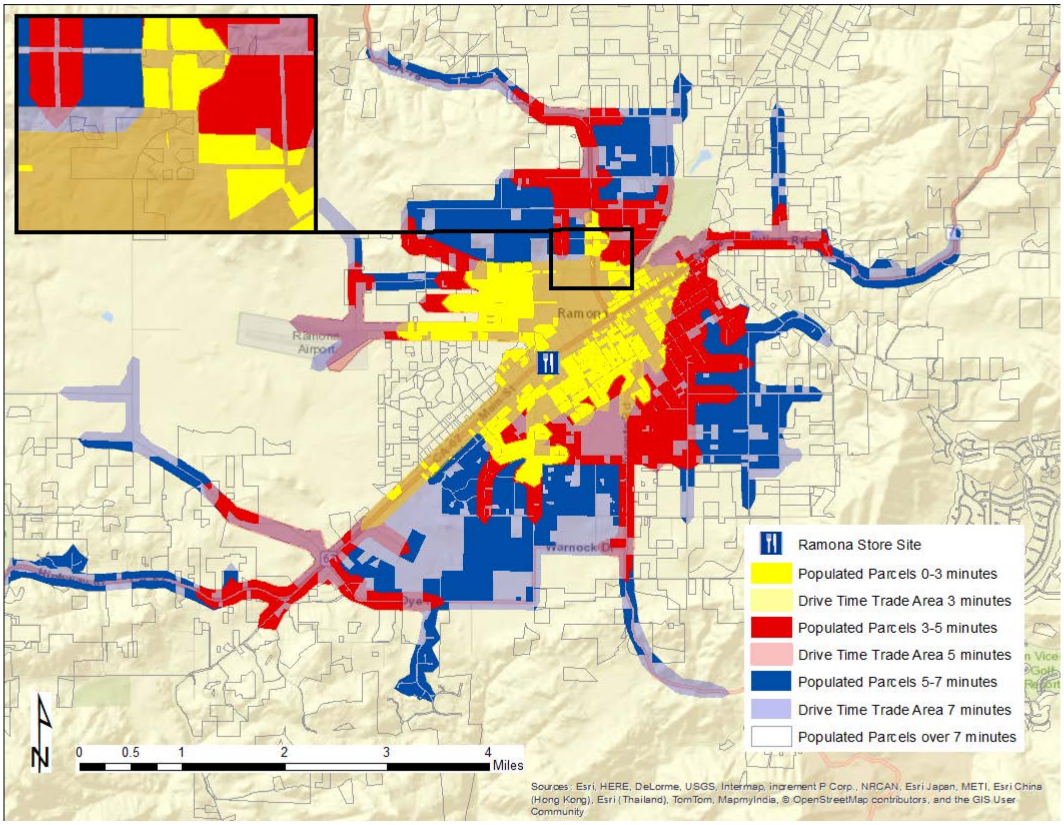

Figure 13: All "Populated parcels" intersecting detailed drive time trade areas of three, five and seven minutes for the Ramona location. These "Populated parcels" are overlaid on a less vibrant detailed drive time trade areas of three, five and seven minutes of the same color. Vibrantly colored areas indicate residences. "Populated parcels" outside the trade area for the site are shown with a gray outline.

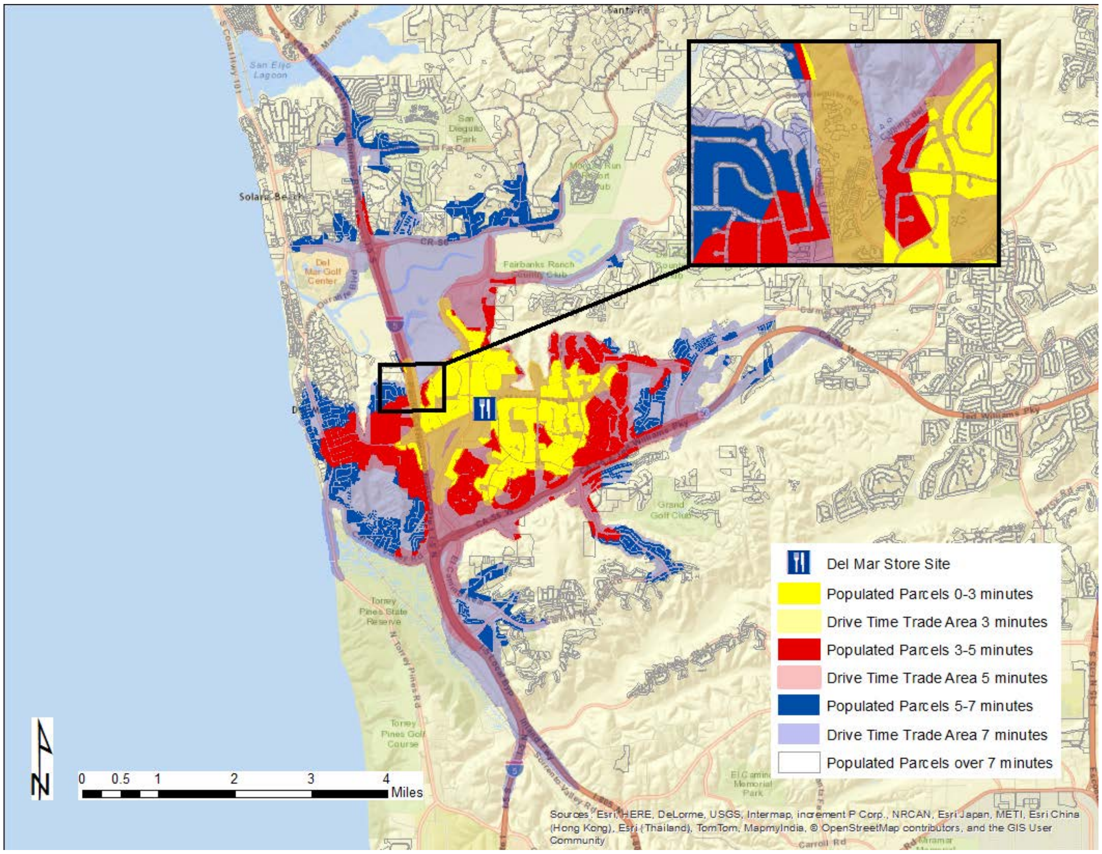

Figure 14 : All "Populated parcels" intersecting detailed drive time trade areas of three, five and seven minutes for the Alpine location. These "Populated parcels" are overlaid on a less vibrant detailed drive time trade areas of three, five and seven minutes of the same color. Vibrantly colored areas indicate where people live. "Populated parcels" outside the trade area for the site are shown with a gray outline.

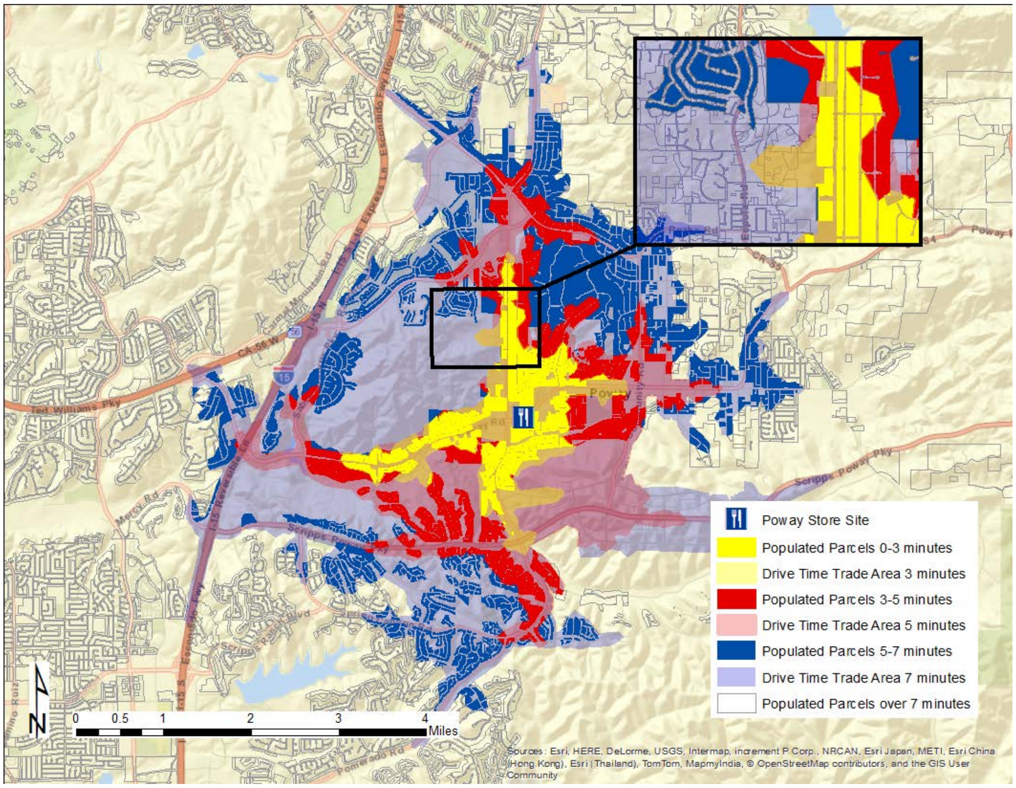

Figure 15: All “Populated parcels” intersecting detailed drive time trade areas of three, five and seven minutes for the Poway location. These “Populated parcels” are overlaid on a less vibrant detailed drive time trade areas of three, five and seven minutes of the same color. Vibrantly colored areas indicate where people live. “Populated parcels” outside the trade area for the site are shown with a gray outline.

The results show that disaggregated polygons (i.e. the Populated parcels) yielded different results for Spatial Overlay of the trade areas than Block Groups and Tracts yielded. Block Groups and Tracts yielded identical Spatial Overlay results. This was due to the weighted block centroid method for retrieval and inclusion in a Spatial Overlay described earlier. Spatial overlays for block groups and tracts would have the same results for the same trade areas given inclusion of their populations based on weighted block centroids.

Table 2 Population Household and associated demographic information for detailed drive time trade areas of store sites at drive times of three, five, and seven minutes. PP is an abbreviation for Populated parcels and BG and T is an abbreviation for Block Groups and Tracts.

Average Household Income												
Site	Drive Time 3 minutes				Drive Time 5 Minutes				Drive Time 7 minutes			
	PP	BG and T	Change	% Change	PP	BG and T	Change	% Change	PP	BG and T	Change	% Change
Alpine	98319	84026	14293	14.5%	99128	82728	16400	16.5%	100920	88908	12012	11.9%
San Diego	142762	144123	-1361	-1.0%	160212	144664	15548	9.7%	166672	151087	15585	9.4%
Poway	114510	93647	20863	18.2%	117737	103389	14348	12.2%	126961	119573	7388	5.8%
Ramona	69902	65531	4371	6.3%	78491	70939	7552	9.6%	83550	79822	3728	4.5%
Per Capita Income												
Site	Drive Time 3 minutes				Drive Time 5 Minutes				Drive Time 7 minutes			
	PP	BG and T	Change	% Change	PP	BG and T	Change	% Change	PP	BG and T	Change	% Change
Alpine	31961	31490	471	1.5%	33432	31406	2026	6.1%	34575	33614	961	2.8%
San Diego	53630	53871	-241	-0.4%	57529	57041	488	0.8%	59022	59161	-139	-0.2%
Poway	32001	30876	1125	3.5%	36223	35004	1219	3.4%	41654	41556	98	0.2%
Ramona	20947	20827	120	0.6%	23925	22717	1208	5.0%	25990	25993	-3	0.0%
Average Net Worth												
Site	Drive Time 3 minutes				Drive Time 5 Minutes				Drive Time 7 minutes			
	PP	BG and T	Change	% Change	PP	BG and T	Change	% Change	PP	BG and T	Change	% Change
Alpine	759178	520392	238786	31.5%	785006	463783	321223	40.9%	823286	582145	241141	29.3%
San Diego	995967	1022112	-26145	-2.6%	1170041	995131	174910	14.9%	1186639	1041376	145263	12.2%
Poway	953819	654552	299267	31.4%	964756	772205	192551	20.0%	1028999	946598	82401	8.0%
Ramona	374462	325255	49207	13.1%	501238	404958	96280	19.2%	578765	536022	42743	7.4%

"Populated parcels" give a more accurate picture of where population is located. Analysis of the San Diego and Poway site Spatial Overlay results show that population counts in suburban areas experience a significant spike when using "Populated parcels." Calculation of demographics using "Populated parcels" for small trade areas yield results similar to block group and tract spatial overlay most notably at higher densities. This is due to the fact that smaller trade areas especially in dense suburban areas are comprised of smaller block groups and tracts which when overlaid yield similar results.

Another trend of note seems to show deviation of results increasing with greater numeric values. Larger percent changes in results seem spurred by larger counts or dollar amounts (i.e. per capita income results show smaller changes than average income, and average income results show smaller changes than average net worth which are numerically larger).

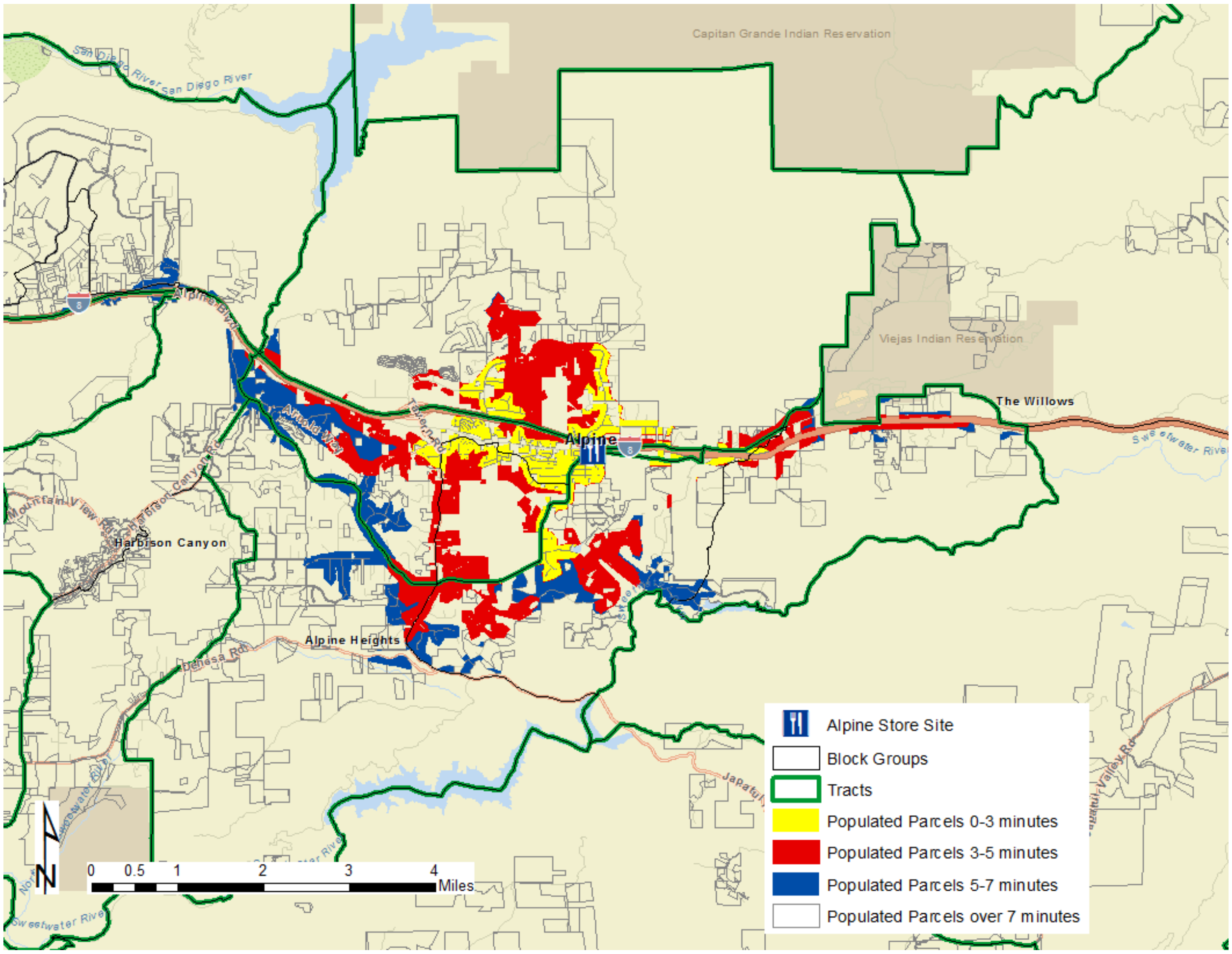

Figure 16 – "Populated parcels" polygons for detailed drive time areas of 3, 5, and 7 minutes. "Populated Parcels" within 3, 5, and 7 minutes are shown in yellow, red, and blue respectively without overlap. Block Groups are outlined in black and Tracts are outlined in green. This figure demonstrates how population distribution is not uniform across census blocks or tracts. The northernmost tract shows population is concentrated mostly in the southern area of the tract.

The results of five, ten, and fifteen minute Drive Time Trade Areas are shown in Table 3. The results at these larger trade areas were also identical for block groups and tracts which is further evidence of the weighted block centroid retrieval. Similar effects for those seen in the results of smaller trade areas were seen in these larger trade areas. Differences in population calculated were found but were not as large relative to smaller trade areas. This makes sense as larger areas intersect an increased number of the census blocks due to their larger size resulting in an averaging effect as more source layer features are include in larger trade areas.

Table 3 Population Household and associated demographic information for detailed drive time trade areas of store sites at drive times of five, ten, and fifteen minutes. PP is an abbreviation for Populated parcels and BG and T is an abbreviation for Block Groups and Tracts.

Population												
Site	Drive Time 3 minutes				Drive Time 5 Minutes				Drive Time 15 minutes			
	PP	BG and T	Change	% Change	PP	BG and T	Change	% Change	PP	BG and T	Change	% Change
Alpine	7631	6056	1575	20.6%	15182	14036	1146	7.5484%	55310	50849	4461	8.0654%
San Diego	30611	28211	2400	7.8%	80954	74182	6772	8.3652%	286593	279359	7234	2.5241%
Ramona	10754	9199	1555	14.5%	17271	15677	1594	9.2293%	25098	22560	2538	10.1124%
Average Household Income												
Site	Drive Time 5 minutes				Drive Time 10 Minutes				Drive Time 15 minutes			
	PP	BG and T	Change	% Change	PP	BG and T	Change	% Change	PP	BG and T	Change	% Change
Alpine	99128	82728	16400	16.5%	105988	95580	10408	9.8%	95246	85572	9674	10.2%
San Diego	160212	144664	15548	9.7%	166152	152141	14011	8.4%	144149	119247	24902	17.3%
Ramona	78491	70939	7552	9.6%	90025	86050	3975	4.4%	97357	94417	2940	3.0%
Per Capita Income												
Site	Drive Time 5 minutes				Drive Time 10 Minutes				Drive Time 15 minutes			
	PP	BG and T	Change	% Change	PP	BG and T	Change	% Change	PP	BG and T	Change	% Change
Alpine	33432	31406	2026	6.1%	36356	35828	528	1.5%	31027	30711	316	1.0%
San Diego	57529	57041	488	0.8%	57592	58867	-1275	-2.2%	46735	46211	524	1.1%
Ramona	23925	22717	1208	5.0%	28723	28001	722	2.5%	31972	31100	872	2.7%
Average Net Worth												
Site	Drive Time 5 minutes				Drive Time 10 Minutes				Drive Time 15 minutes			
	PP	BG and T	Change	% Change	PP	BG and T	Change	% Change	PP	BG and T	Change	% Change
Alpine	785006	463783	321223	40.9%	934549	750152	184397	19.7%	840523	682198	158325	18.8%
San Diego	1170041	995131	174910	14.9%	1188120	1059119	129001	10.9%	1029372	786510	242862	23.6%
Ramona	501238	404958	96280	19.2%	668838	612131	56707	8.5%	792018	762617	29401	3.7%

4.2 Comparison of Trade Areas Created by Distance Measures to Drive Time Trade Area

The standard method of calculating population and demographic information used in real estate marketing and decision making as described in Chapter 1 is to utilize radial distance rings in such analysis. This case study also sought to determine how calculated catchment population changed using radial distance and two other trade area creation methods using network distance measures as well as drive time distance measures in miles and minutes respectively. Calculations for each of the four store sites in San Diego were conducted at 3, 5, and 7 mile radial distance, and network (road) distance as well as 3, 5, and 7 minute drive times. A spatial overlay for all of these differently defined trade areas with both the “Populated parcels” and population aggregates was conducted to find the effect on the results. Figure 18 below pictures these differently defined trade areas for each store at 7 minute and mile distance measures as an example.

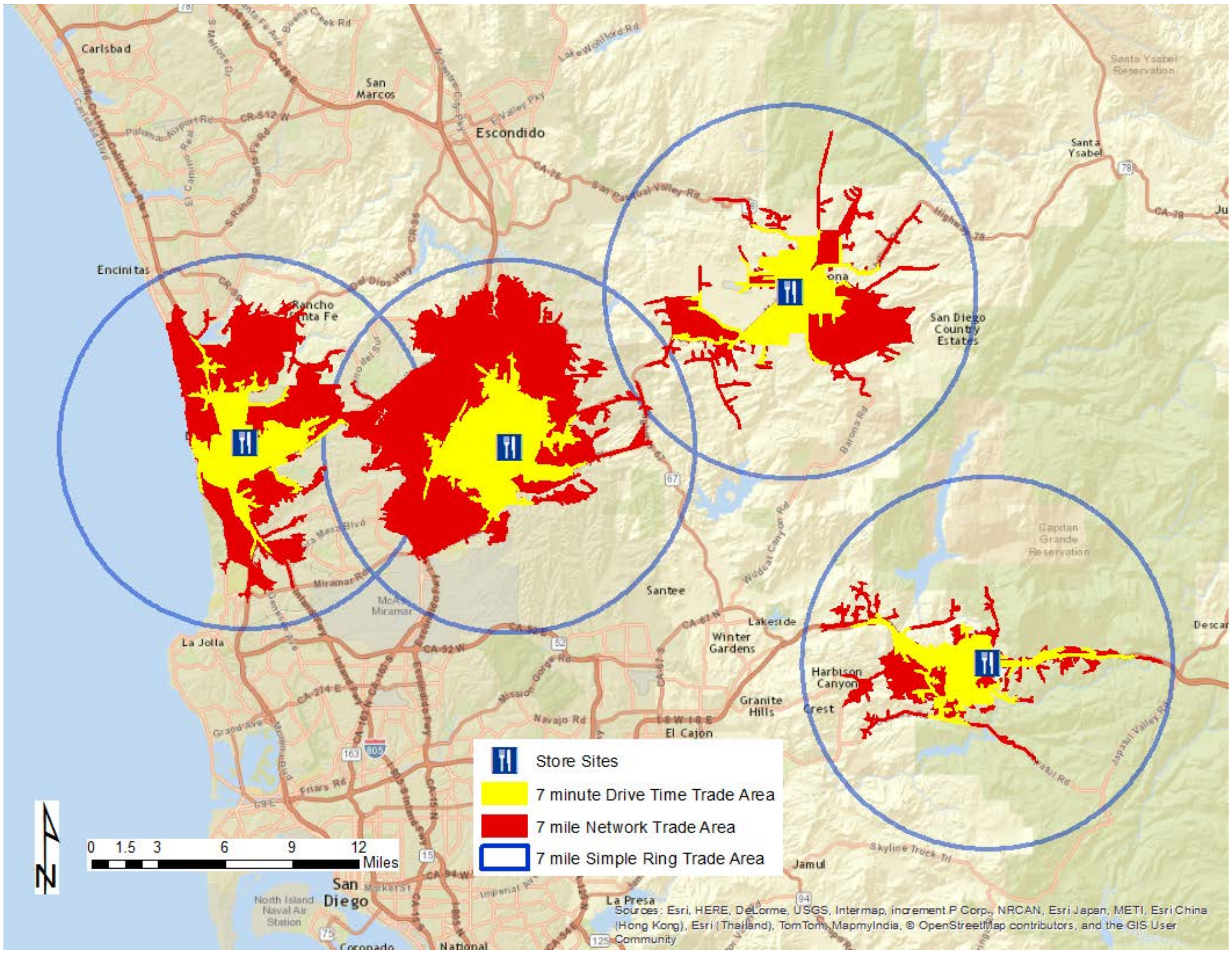

Figure 17: 7 minute Drive Time Trade Areas and 7 mile network and radial distance trade areas.

Table 4 shows the results calculated for “Populated parcels” and census aggregates for 3 minute drive time trade areas, beneath these are the results for “Populated parcels” and census aggregates for 3 mile network distance trade areas, with the results for “Populated parcels” and census aggregates for 3 mile radial trade areas at the bottom of the table. Table 5 and Table 6 show similar results but at distance measures of 5 minutes and miles and 7 minutes and miles respectively. The results of these analyses are discussed below.

Table 4 Population Household and associated demographic information Alpine, San Diego, Poway, and Ramona sites for all trade areas of distance 3. PP is an abbreviation for Populated parcels and CENSUS is short for census aggregates. POP is an abbreviation for Population. AVGHHINC is an abbreviation for average household income. PCI is an abbreviation for per capita income. AVGNW is an abbreviation for average net worth.

3					
Drive Time Trade Area Minutes	**SITE**	**POP PP**	**AVGHHINC PP**	**PCI PP**	**AVGNW PP**
	Alpine	3694	85096	31961	530545
	San Diego	15023	141760	53630	972439
	Poway	9753	96211	32001	669394
	Ramona	7418	65560	20947	316226
	SITE	**POP CENSUS**	**AVGHHINC CENSUS**	**PCI CENSUS**	**AVGNW CENSUS**
	Alpine	3642	84026	31490	520392
	San Diego	12372	144123	53871	1022112
	Poway	7330	93647	30876	654552
	Ramona	7148	65531	20827	325255
Network Distance Trade Area Miles	**SITE**	**POP PP**	**AVGHHINC PP**	**PCI PP**	**AVGNW PP**
	Alpine	9648	94560	35259	690459
	San Diego	39910	156907	59431	1109670
	Poway	43062	116033	38761	927667
	Ramona	12810	76545	24668	471347
	SITE	**POP CENSUS**	**AVGHHINC CENSUS**	**PCI CENSUS**	**AVGNW CENSUS**
	Alpine	7481	90697	34065	615426
	San Diego	37205	149484	58776	1027426
	Poway	38810	113081	38412	901039
	Ramona	11897	76539	24654	477058
Simple Ring Trade Area Miles	**SITE**	**POP PP**	**AVGHHINC PP**	**PCI PP**	**AVGNW PP**
	Alpine	13511	99248	36689	789520
	San Diego	66172	159157	59366	1096045
	Poway	85962	120734	41597	948290
	Ramona	18653	83108	27196	565243
	SITE	**POP CENSUS**	**AVGHHINC CENSUS**	**PCI CENSUS**	**AVGNW CENSUS**
	Alpine	13401	98498	36708	773433
	San Diego	66464	155702	59518	1059901
	Poway	85422	117977	41433	906868
	Ramona	18349	83934	27174	578919

The results in Table 4 show that for 3 minute and mile distance measures calculations of population and demographics using the same trade area were roughly the same when using either the “Populated parcels” or census aggregates. The differences that do exist show that for each population and demographic characteristic calculations using the “Populated parcels” are higher for each result as opposed to those calculations using census aggregates. These differences in results are more pronounced in urban store trade areas than the rural store trade areas.

While comparison of the results found using the same trade areas had little change, large changes can be seen in the results found using the different trade areas. As can be expected using radial distance provides the highest estimation of population with figures almost double in some cases the population of the network distance trade areas and roughly four times the population of drive time trade areas. This same trend is however not seen in the results of the demographics as demographics will likely be similar at shorter distances due to autocorrelation.

Table 5 shows results for trade areas defined using 5 minute and mile distance measures. The results for trade areas at these distances are similar to those found at 3 minute and mile distance measures. Generally calculations of demographics were similar but tend to be underestimated using aggregated population when compared to those found using “Populated parcels.” Population calculations for trade areas found at distance measures of 5 follow the same trends as those found at trade areas found at distance measures of 3. The one exception would be Store 2 which is coastal and therefore trade areas are constrained by the coast and limited in expansion to one side. As a result the differences in population are not to the same magnitude for the coastal store site as the other store sites.

Table 5 Population Household and associated demographic information Alpine, San Diego, Poway, and Ramona sites for all trade areas of distance 5. PP is an abbreviation for Populated parcels and CENSUS is short for census aggregates. POP is an abbreviation for Population. AVGHHINC is an abbreviation for average household income. PCI is an abbreviation for per capita income. AVGNW is an abbreviation for average net worth.

5					
Drive Time Trade Area Minutes	**SITE**	**POP PP**	**AVGHHINC PP**	**PCI PP**	**AVGNW PP**
	Alpine	7631	88916	33432	584622
	San Diego	30611	151948	57529	1072966
	Poway	25212	109135	36223	839904
	Ramona	10754	74152	23925	440390
	SITE	**POP CENSUS**	**AVGHHINC CENSUS**	**PCI CENSUS**	**AVGNW CENSUS**
	Alpine	6056	82728	31406	463783
	San Diego	28211	144664	57041	995131
	Poway	20870	103389	35004	772205
	Ramona	9199	70939	22717	404958
Network Distance Trade Area Miles	**SITE**	**POP PP**	**AVGHHINC PP**	**PCI PP**	**AVGNW PP**
	Alpine	13353	98960	36423	781709
	San Diego	69787	156715	59065	1095429
	Poway	113522	121273	42503	962195
	Ramona	16494	83792	27647	584441
	SITE	**POP CENSUS**	**AVGHHINC CENSUS**	**PCI CENSUS**	**AVGNW CENSUS**
	Alpine	12052	96254	35900	732027
	San Diego	65994	151621	59265	1052956
	Poway	108880	117463	42151	916911
	Ramona	14962	82791	26915	573707
Simple Ring Trade Area Miles	**SITE**	**POP PP**	**AVGHHINC PP**	**PCI PP**	**AVGNW PP**
	Alpine	18156	102501	37433	847659
	San Diego	105899	154838	56715	1116241
	Poway	189917	123234	42483	991728
	Ramona	29511	95773	32216	772277
	SITE	**POP CENSUS**	**AVGHHINC CENSUS**	**PCI CENSUS**	**AVGNW CENSUS**
	Alpine	18001	101710	37474	832697
	San Diego	106595	150448	56524	1059742
	Poway	187782	119883	42401	954646
	Ramona	29604	96957	32246	794279

Table 6 shows results for trade areas defined using 7 minute and mile distance measures to construct trade areas. The results are similar to those results found using trade areas delineated with 3 and 5 minute and mile distance measures. As with the previously defined smaller trade areas the 7 minute and mile trade areas reflect minimal change in population calculations within the same trade area but large population changes found using different distance measures for trade area delineation. This difference is however not as drastic in the larger distance trade area population calculations. However the smaller distance trade areas were more uniform in their results for demographics. At larger distance measure more room for deviation is more likely which is reflected in the results having less consistent trending than the smaller distance trade areas.

Table 6 Population Household and associated demographic information Alpine, San Diego, Poway, and Ramona sites for all trade areas of distance 7. PP is an abbreviation for Populated parcels and CENSUS is short for census aggregates. POP is an abbreviation for Population. AVGHHINC is an abbreviation for average household income. PCI is an abbreviation for per capita income. AVGNW is an abbreviation for average net worth

7					
Drive Time Trade Area Minutes	SITE	POP PP	AVGHHINC PP	PCI PP	AVGNW PP
	Alpine	9305	92624	34575	657480
	San Diego	45494	156598	59022	1096948
	Poway	61192	121309	41654	964233
	Ramona	13955	79535	25990	523983
	SITE	POP CENSUS	AVGHHINC CENSUS	PCI CENSUS	AVGNW CENSUS
	Alpine	7595	88908	33614	582145
	San Diego	40229	151087	59161	1041376
	Poway	55983	119573	41556	946598
	Ramona	13105	79822	25993	536022
Network Distance Trade Area Miles	SITE	POP PP	AVGHHINC PP	PCI PP	AVGNW PP
	Alpine	16445	100424	36843	818781
	San Diego	87348	154452	58193	1078976
	Poway	222146	121886	42246	966998
	Ramona	23348	94043	31246	745614
	SITE	POP CENSUS	AVGHHINC CENSUS	PCI CENSUS	AVGNW CENSUS
	Alpine	14731	97190	36273	767085
	San Diego	81820	152207	59113	1056063
	Poway	216912	118648	42187	936064
	Ramona	21594	94123	30805	750308
Simple Ring Trade Area Miles	SITE	POP PP	AVGHHINC PP	PCI PP	AVGNW PP
	Alpine	26544	103864	38000	889477
	San Diego	273615	125537	45920	882579
	Poway	291443	121797	41687	958207
	Ramona	35503	101206	34140	851466
	SITE	POP CENSUS	AVGHHINC CENSUS	PCI CENSUS	AVGNW CENSUS
	Alpine	25939	102508	38019	871014
	San Diego	284790	123246	45456	839547
	Poway	294964	119761	41589	937457
	Ramona	35697	102002	34160	870044

4.3 Summary of Results

Census population aggregates produced calculations of Drive Time Trade Area catchment population which were less than calculations produced by “Populated Parcels”. Population differences were more significant at smaller distances but insignificant overall. Significant differences in catchment population were found when different distance measures were used for creation of trade areas. Simple Ring trade area estimates were roughly double and quadruple population counts than those found using network distance and drive time distances.

Simple ring is most common trade area creation method used in retail real estate marketing and subsequent decision making. The large differences in population estimates show that radial distance measures overestimate catchment population and how crucial trade area definition is to catchment population calculations. The differences found in this case study show further investigations into trade area delineation and disaggregating population are warranted. Recommendations for further studies are given in the next Chapter.

CHAPTER 5: DISCUSSION AND CONCLUSIONS

The results of this study expose possible shortcomings of methods of analyses currently used in retail real estate marketing and decision making. This case study explored how calculation of catchment population and related demographics using “Populated parcels” contrasted with analysis using census aggregates commonly used in real estate. It also examined how more precise methods for calculating a store’s trade area affected calculated catchment population and related demographics compared with catchment population calculated using simple ring and road network distance trade areas, typically used in real estate.

Concentrating population data to areas where population lives provides improved calculations of catchment population compared to those calculated using census aggregates. When used to calculate catchment population and corresponding characteristics for Drive Time trade areas, census aggregates only include population where trade areas pass through the block centroids within the census aggregates. As such, all population for those census blocks whose centroids are not within the trade area are fully omitted from catchment population calculations despite having some population within a trade area. Simultaneously, census blocks whose centroids are within a trade area have the entire population of the census blocks included in catchment population calculations for trade areas.

Concentrating population to only residential areas and weighting those areas by percentage of their area as a ratio of total residential areas within their census block provides a more precise estimate of population. This overcomes the shortcoming of total inclusion or total exclusion of population using weighted block centroid inclusion for calculation. Inclusion of previously omitted population and exclusion of some portions of previously over-included population aggregates results in more accurate catchment population calculations which were more inclusive of actual population within a trade area. Results found using census block groups and tracts produced identical results which reinforces this idea. Block groups and tracts should ideally have results that vary as they are not identical. Results for both census units should intersect each aggregate level differently, pointing to the problem with using the aggregate centroids as the qualification for inclusion in calculated characteristics is a flaw in this approach.

This study also showed how using “Populated parcels” produces different calculations of catchment population than census aggregates. Calculated catchment population differences were minimal, however, the calculated related demographics experienced much larger percent changes. Even if catchment population estimates are similar, large differences in demographics show the potential for problems in subsequent real estate modeling or decision making using these demographics for analysis. These differences in population and demographic calculations found with concentrated population vary depending upon population density of the trade area and other study area community characteristics.

While this study showed how using concentrated population data may provide greater insight, obtaining population characteristics collected via surveys at the block level is also necessary as much demographic information is not distributed below Block Group aggregates. This would be very time consuming and add cost. However, this study’s results display ecological fallacy, where characteristics of larger aggregate data were imposed on disaggregated data and may not correlate with the new lower level geography. Using parcel level survey data would allow for characteristics to be verified against new lower level polygon aggregates and disaggregates.

The effects of using concentrated population and aggregate population data were then extended to using different distance trade area creation methods as well. Distance measures used were road network minutes and miles as well as simple ring radial miles which is most commonly used in real estate to construct trade areas. Distances of 3, 5, and 7 minutes and miles were used to construct store trade areas for each site and catchment population was calculated using each set of population data.

The results of this investigation showed similar results to the previous results for all trade areas. However, what was most notable were the drastic increases in trade area size and calculated catchment population from trade areas constructed using network minutes to miles and again from network miles to simple ring trade areas. Population calculations for network minutes to network miles roughly doubled. The same effects can be seen from network miles to simple ring miles, resulting in roughly four times the calculated population from network minutes to radial miles.

The shortcomings of using simple rings or radial distance measures were discussed in Chapter two. Network distance is a much better measure of distance traveled and Drive Time distance measured in minutes is an even better indicator of actual costs of reaching a store site which impacts a consumer’s willingness to travel. Despite this, simple ring distances are commonly used in real estate marketing and decision making. The large differences in calculations of catchment population using these different trade area creation approaches show what may be a major shortcoming of using simple rings for real estate marketing and analysis as simple rings largely overestimate catchment population.

The investigations undertaken in this case study show that using concentrated data may result in large differences in demographic calculations as opposed to using census aggregates. Additionally it was shown that traditional methods of trade area calculation may drastically overestimate population in a trade area, meaning that the actual population likely to patronize a store may be less by roughly double or four times that calculated traditionally. These shortcomings can be detrimental to real estate decision making and ultimately could cost a great deal of money to investors, retail business owners, and banks which finance them a great deal of money. Further investigation into these outcomes is warranted and future studies could be expanded in a variety of ways.

5.1 Directions for further research

This study shows that further investigation of how dasymetric mapping along with different trade area creation methods impact a store site's calculated catchment population and demographics is warranted. Possible directions for future research should include augmenting this study with new methods of weighting population in dasymetric mapping or using other trade area creation techniques to define trade areas to see these effects. These and other possibilities for future research are discussed in the following sections.

5.1.1 Weighting Population

Population was areally weighted by percentage of area for all residential parcels within a block. Other methods for weighting population could also be used in future analysis and again verified with survey data to see how these methods affect outcomes of disaggregating populations and weighting the corresponding characteristics. Three different weighting methods include weighting residential land areas by number of units from assessor data, weighting population by different land use classifications, and weighting populations for all land uses by land use as well as area.

A simple improvement to the current study's weighting would be to weight the value by number of units in the case of multifamily and student housing as well as by number of APN's in the case of condominiums. This would ensure that each area would be weighted by possible households. Average Household Size would be multiplied against this weighting to get population estimates. Further investigations could also be done which take in to account the vacancy factor for the area homes which could be found from real estate literature derived from surveys. Without taking into account a vacancy rate, an area could be overestimated, though with a current vacancy factor of two percent in San Diego County overestimation may not be significant in this study.

Weighting land uses by land use classification means weighting all existing built structures by land use classification. This would be done to simulate where people are during the day. For example, residential and commercial properties could have a higher weight for their classifications than say industrial as industrial properties tend to have few workers. Retail parcels would be weighted similar to industrial for number of employees but have an increased weighting due to customer draw, which would increase likelihood that people would be coming from these places.

An extension of weighting land use by land use classification would be to weight land uses by land use and area as well. Weighting by area would give increased weight to larger areas. Weighting by a land use classification alone would mean that larger residential areas would be weighted the same as small residential areas. By weighting populated areas by both land use classification and area as weights areas more likely to have population as determined by land use would also be weighted proportionally to their size.

5.1.2 Different Trade Area Techniques

Another variation on the current study would be to use the disaggregated data to see how different trade area creation methods affect the resulting population of trade areas for store site's trade area population, demographic and economic characteristics. This method would use different trade area creation methods in Business Analyst not utilized in this study. For example, Thiessen polygons could be used to construct trade areas defined by creating trade areas where a store's trade area is defined by all points nearest to the store than any of its competitors.

Another trade area method where store competitors are considered in defining a store's trade area is the Huff Gravity Model. In this model, store sites are assigned trade areas by many factors similar to the regression method discussed in Chapter two. Other factors such as store square footage, which is often used as a proxy for store sales, might also be used in the model.

5.1.3 Other possible improvements for the future

Additional improvements to this study would involve doing more store trade areas for more sites and in different counties with varied population densities. Studies similar to this investigation could be conducted for other suburban counties in southern California and other urban counties in areas like Chicago and very rural counties in states like Arizona or Wyoming on the opposite end of the spectrum. Having a larger sample size of sites of varying densities would allow for greater identification of trends resulting from different site population densities.

Such additional methods would be important to investigate because retail sites could fail or succeed depending on estimates of catchment population and demographics. Better information would increase level of success of site suitability analysis and allow financers to be more accurate. Real estate transactions and sale lease backs are based on these metrics and priced accordingly. Improper or less accurate data can lead to poor investing and cost real people real money.

5.2 Conclusions

The methodology used in this case study produced results which were more representative of the population within each trade area than the calculations found with the census aggregates. Removing portions of census aggregates and concentrating population to where people actually live is an improvement over the centroid inclusion method which does not take into account actual population distribution. This case study shows that further investigation into this type of population concentration is warranted. Future directions for improvement on this research described earlier in this chapter and analysis of these effects will lead to a far better calculation of catchment population and population characteristics and could be immensely beneficial to calculations of store site suitability analysis. Employing these techniques and improvements could also be a differentiating factor for a researcher or real estate agent in marketing oneself versus their competitors.

Regardless of the methods used, calculation of underlying population and population characteristics should be as accurate as possible. Businesses that have better information are better equipped to pick a location for a store that will help to ensure the success of the store and of the investment in the store. Additionally in real estate sales, retail property owners, banks and investors make determinations about store profitability and worth based on many of the population characteristics determined by researchers for each property. Information of better quality and precision can lead to better decision making and increase the probability of a site's success and of investor and bank returns.

BIBLIOGRAPHY

Amaral, Silviana, Andre Augusto Gavlak, Maria Isabel Sobral Escada, Montiero, and Anotnio Miguel Vieira. 2012. "Using Remoe Sensing and Tract Data to Improve Representation of Population Spatial Distribution: Case Studies in the Brazilian Amazon." *Population and Environment* 34: 142-170.

Birkin, Mark, Graham Clarke, Martin Clarke, and Alan Wilson. 1996. *Intelligent GIS : Location Decisions and Strategic Planning*. 1st ed. Vol. 1. New York, NY: Pearson Professional Limited.

Fenker, Richard and Juli Zoota. 2001. "Intuitive Retail Modelling: Does Science have Anything to Offer?" *Journal of Corp Real Estate* 3 (3): 248-259.

Ignizio, Drew A. and Paul A. Zandbergen. 2010. "Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates." *Cartography and Geographic Information Science* 37: 199-214.

Kelly, Jason P. 2012. "The Strategic use of Prisons in Partisan Gerrymandering." *Legislative Studies Quarterly* 37 (1): 117-134.

Mennis, Jeremy. 2003. "Generating Surface Models of Population using Dasymetric Mapping." *Professional Geographer* 55 (1): 31-42.

Miller, Fred L. 2010. *Getting to Know ESRI Business Analyst*. 380 New York Street, Redlands, CA: Esri Press.

Mitchell, Andy. 2009. *The Esri Guide to GIS Analysis Volume 2 :Spatial Measurement and Statistics*. First ed. Redlands, CA: Esri Press.

Peters, Alan and Heather MacDonald. 2004. *Unlocking the Census with GIS*. First ed. Redlands, CA: ESRI Press.

Qiu, Fang and Robert Cromley. 2013. "Areal Interpolation and Dasymetric Modeling." *Geographical Analysis* 45 (3): 213-215.

Tobler, W. R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46 (Supplement: Proceedings. International Geographical Union. Commission on Quantitative Methods): 234-240.

Wen, Xingping and Hu Guangdu Yang Xiaofang. 2011.*Journal of Indian Science of Remote Sensing* June 2011 (39): 193-201.

Wong, David. 2009 "The Modifiable Areal Unit Problem (MAUP)." In *The SAGE Handbook of Spatial Analysis*, edited by A. Stewart Fotheringham and Peter A. Rogerson, 105-125. London: SAGE Publications Ltd.

Zandbergen, Paul A. 2011. "Dasymetric Mapping using High Resolution Address Point Datasets." *Transactions in GIS* 15: 5-27.