Signature of Thesis Supervisor

Author

Helmut Brückler, BSc

Development of a Custom Site Selection Model for Microbreweries in San Diego County based on Geodemographics and Twitter Data

Master Thesis

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur (Master of Science)

Degree Program

Geodesy and Geomatics Engineering (066 421)

Vienna University of Technology
Department of Geoinformation

Supervisors:

O.Univ.Prof. Andrew Frank (Vienna University of Technology)

Prof. Piotr Jankowski (San Diego State University)

San Diego, 2014

Statutory Declaration

I declare in lieu of an oath that I have written this master thesis myself and that I have not used any sources or resources other than stated for its preparation. This master thesis has not been submitted elsewhere for examination purposes.

Vienna, on --

Acknowledgements

First of all I want to thank my family for the great support throughout my entire academic studies and also friends and colleagues who always had an open ear for me.

I also want to express a big thank you to Professor Piotr Jankowski - my advisor at San Diego State University - for giving me the opportunity to work on my master research project at his department and his extensive support throughout my stay. A lot of credit also goes to Prof. Andrew Frank for supervising my master thesis at the Vienna University of Technology.

Furthermore I appreciate the guidance of Angelika Schweighart through the Marshall Plan Scholarship application process.

And finally I want pay tribute to the Marshall Plan Foundation for supporting me financially and opening the door to an unforgettable experience!

"Everything is theoretically impossible, until it is done."
(Robert A. Heinlein)

Abstract

The performance of a microbrewery is heavily dependent on its location. Badly chosen locations, often due to ad-hoc or poorly conducted site selection analysis, are a key factor for businesses to go down. The emergence and rapid development of Geographic Information Systems (GIS) with their spatial analytical capabilities open new doors to strategic site performance analysis. Combining GIS analysis and mathematical models with geodemographic and social media data can help planners and decision makers in the essential evaluation process of potential candidate sites. The focus of this research project lies on gaining new knowledge about craft beer consumption characteristics and behaviors, and utilizing this knowledge to develop of a custom site selection model for microbreweries which calculates the number of potential craft beer consumers within a specified geographic area. The study area comprises the entire County of San Diego. An exploratory analysis is conducted in order to gain new knowledge of craft beer consumers based on a set of Twitter data collected over a time period of one year. All tweets contain locational, temporal and textual data on which spatial, qualitative and quantitative analyses methods are applied to obtain information about the spatial distribution and demographic characteristics of craft beer consumers. The demographic information gained from the exploratory analysis is used for the development of the site selection model which corresponds to a geodemographic probability model taking as an input the demographic characteristics of a specific geographical area. The model is applied in two location analysis scenarios at different scales. The first one is a macro level analysis with the objective of estimating the number of potential craft beer consumers for each ZIP code in San Diego County. The second scenario corresponds to a micro level site suitability analysis, evaluating the feasibility for opening a new microbrewery on a certain number of vacant sites which are either for sale or rent.

Kurzfassung

Der wirtschaftliche Erfolg von Kleinbrauereien ist sehr stark von dessen Standort abhängig. Fehlende oder nur oberflächlich durchgeführte Standortanalysen sind oft die Hauptursache für das Nichterreichen der erwarteten Gäste- und Umsatzzahlen. Der Einsatz von Geographischen Informationssystemen (GIS) ermöglicht es, potentielle Standorte sowohl räumlich als auch attributiv zu analysieren und daraus folgend auf dessen Tauglichkeit hinsichtlich des beabsichtigen Geschäftszweckes zu schließen. Die Kombination von GIS Analysen und mathematischen Modellen mit geodemographischen und Social Media Daten kann Entscheidungsträger im Evaluierungsprozess von potentiellen Standorten unterstützen. Ziel dieses Forschungsprojektes ist die Gewinnung neuer Informationen über die Charakteristiken von Craft-Bier Konsumenten und darauf stützend die Entwicklung eines Standortanalysemodels welches die Anzahl der potentiellen Craft-Bier Konsumenten für ein bestimmtes Gebiet approximiert. Als Forschungsgebiet wurde die amerikanische County San Diego, Kalifornien gewählt. Zuerst wird eine explorative Datenanalyse am vorliegenden Konsumentendatenbestand durchgeführt um Kenntnisse über „typische“ Craft-Bier Konsumenten zu erlangen. Der Datenbestand entspricht einem georeferenzierten Twitter Datensatz mit Beziehung zu Craft-Bier mit einem Beobachtungszeitraum von einem Jahr. Qualitative, Quantitative und räumliche Analyseverfahren werden angewandt um die entsprechenden Konsumenteninformationen zu gewinnen. Im zweiten Schritt wird ein Standortanalysemodell implementiert um einerseits die Anzahl der Craft-Bier Konsumenten innerhalb einer Region zu schätzen und anderseits das Einzugsgebiet einzelner potentieller Standorte räumlich zu bestimmen.

Table of Contents

1 Introduction

1.1 Motivation

Selecting the optimum location for a new business, especially in retail and gastronomy, with the objective of maximizing the profitability of the proposed business intent most often involves comprehensive and complex analysis of potential market areas on both macro- and micro level scale. Combining geodemographic, local market and socioeconomic information with GIS analysis and mathematical modeling techniques can help planners and decision makers in the evaluation process of potential candidate sites.

The performance of a retail and gastronomy business is mainly influenced by two main factors. This is on the one hand, the profile of the site itself and on the other hand the characteristics of its inter-urban surroundings. The former refers to criteria such as store attractiveness, size, popularity, accessibility, drive time to the store or parking options and the latter refers to factors such as geodemographic structures, market potential, customer flows, consumer demand and spending habits, or the spatial relation to complementary and competitor facilities. Some of these performance relevant criteria are related to each other and hence quantifying those attributes and their relationship enables the construction of powerful models for spatial analysis helping planners and decision makers to approximate the delineation of the market area of a potential candidate location or derive an estimate of the stores profitability serving as the decision criterion whether a candidate site is feasible for a proposed business intent.

In this research project an automated custom site selection model for microbreweries is developed which is based on geodemographic and social media data. The study area comprises the bounding box of San Diego County. Approximately 70 existing micro breweries are located within the county boundaries, making it one of the counties with the most openings over the last 40 years and hence San Diego is often referred to as the craft beer capital of the US. A microbrewery is a brewery which is typically much smaller than large-scale corporate breweries and their products are generally characterized by their emphasis on quality, flavor and brewing technique (Boteler 2009). The beer products distributed by these companies are referred to as craft beer. Due to the difference in flavor and the higher sales price of this type of beer, the clientele consuming craft beer differs from the one consuming mass-production beer. The demographic characteristics of craft beer consumers need to be examined in detail in order to make feasible site suitability assessments.

Why does the location of a new microbrewery matter?

Another important difference to mass-production beer companies is marketing related. Mass-production beer companies produce their products in large factories. Typically these products are not sold at the same location where they are processed, but are placed in distribution channels and usually purchased in stores or bars. Hence the location of a mass production beer brewery is irrelevant in terms of the demographic characteristics of the people living nearby. On the other hand, microbreweries go along with the tradition of both brewing and selling their product at the same place. At so called on-premise brewpubs customers can purchase the brewed craft beer and most of these breweries also offer dining opportunities. Additionally, some microbreweries also have tasting rooms where people can sample the craft beer collection of the brewery, which helps them to acquire new customers. Therefore the characteristics of the people that live within the surroundings of a microbrewery have a huge impact on the performance of microbrew businesses. The location factor is very crucial and the geodemographic characteristics of the people living within the catchment area of the candidate site need to be examined extensively in order to be profitable.

The rise of social media and its value in knowledge mining

The rise of social media networks such as Twitter, Facebook or Foursquare has led to a massive accumulation of various types of data which has been generated 'voluntarily' by users in forms of web blogs, podcasts or micro-blogging. Social Media can be defined as any form of electronic communication through which users create online communities to share information, ideas, personal messages, and other content (Merriam-Webster 2004). This data created by users of these online communities can be a very valuable source for analysis purposes since it reveals information about people's characteristics and behaviors regarding certain activities. Analysis on social media data can be conducted to gain new knowledge about the insights of consumers such as their demographic characteristics, what products they consume and how frequently they visit a restaurant or brew pub. This information can then be incorporated in site suitability analysis workflows to enhance the decision making process in site selection. The availability of data where users disclose their posts, referring to locational information, can have an additional value to marketing analysis since it reveals knowledge on how consumers locomote through space, where most of the consumers are concentrated and what products they consume where and at what time. Such spatiotemporal data can reflect the activity patterns of users (Noulas et el, 2011) and can be used to evaluate if social media users are close to each other geographically, or to what extent the virtual network space resembles the geographical space (Kulshrestha et al. 2012; Takhteyev et al. 2012). The ability to determine the residential location of consumers enables planners and decision makers to assess the catchment area of an existing microbrewery, referring to the area around the microbrewery where the vast majority of the customers originate from. Determining the delineation of such catchment areas and analyzing the demographic characteristics within these areas can generate valuable knowledge which can then be used for future site selection projects in the process of evaluating the feasibility of each candidate site location regarding their catchment areas. At this point of time, the usage of social media data for analysis purposes especially for site suitability analysis is still a rather novel approach (see chapter 2 – Literature Review) and hence entails new opportunities and challenges on knowledge mining for site selection models.

Geodemographics – the characteristics of people based on where they live

In the process of selecting a new site for a microbrewery, examining the geodemographic characteristics of the people that live within the main catchment area of each candidate site is an inevitable task in approximating a realistic score of the number of potential customers for each site. Geodemographics is the analysis of people by where they live (Sleight 2007) and can be labeled as the study of population types and their dynamics as they vary by geographical area (Birkin et al 1998). Knowledge of the demographic profile of a typical craft beer consumer is required in order to make plausible assessments of whether a candidate site is feasible regarding the potential clientele living in vicinity of the particular site. People are characterized by specific demographic variables such as age, gender or race. Different approaches and methods can be applied to combine these variables to retrieve a joint statistic for assessing the feasibility of a candidate site regarding the quantity and quality of potential customers such as geodemographic segmentation or geodemographic probability models. These various approaches will be discussed in more detail in chapter 2.

1.2 Background

This chapter provides explanations of the theoretical background of the core areas of this research. It starts out with a brief summary of the US micro brew industry including historic development, market overview, consumer information and trends. In a subsequent subsection, the social media network Twitter will be covered in detail since the exploratory analysis conducted in chapter 3 will be based on data retrieved from Twitter. Furthermore, the theoretical basis of geodemographics will be explained and the chapter closes with an overview of the most prominent models used in site selection.

1.2.1 Micro Breweries in the US

The microbrew industry is one of the few industries in the US which had undergone a constant growth in turnover, products manufactured, and new openings (see Figures 1 and 2).

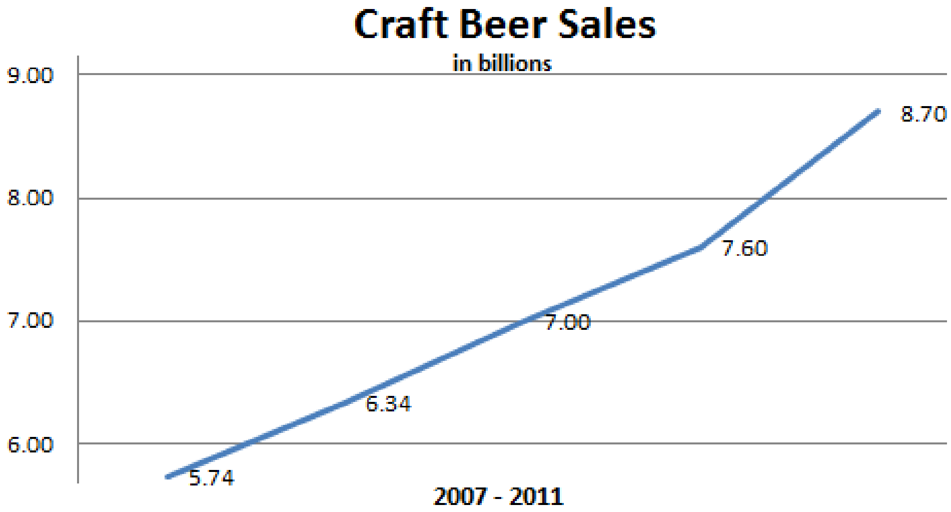


Figure 1: *Craft beer sales from 2007 – 2011 in the US* ***(Brewers Association).***

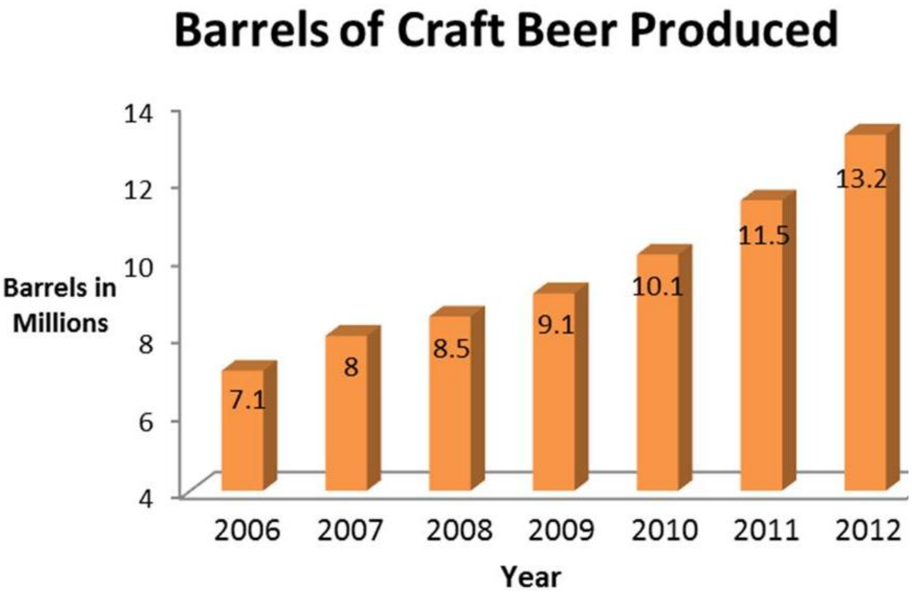


Figure 2: *Barrels of beer produced in the US from 2006 to 2012* ***(Brewers Association).***

The popularity of premium craft beer has been growing since the microbrewery boom in the early 1970s. The microbrewery movement began on the west coast, where Fritz Maytag opened Anchor Brewing Company in 1965, which is now considered the first craft brewery in the US (Lapoint, 2012). The United States experienced a large increase in the availability of beer and a sharp increase in the number of microbreweries due to the legalization of brew pubs in 1978 (Snyder 2012). Many homebrewers looked for ways in which they could brew high quality beer and share or sell it to larger markets than just their friends (Lapoint, 2012). Between 1990 and 2010, 1,376 microbreweries were established in the United States. Rapid expansion in the 1980's led to an oversaturated market and volatility peaked in 1996 when 333 microbreweries opened, and only 36 closed. The volatile growth of the microbrewery industry has been compared to the dot-com boom in the '90's (Snydor 2012). Especially during the years of 2006 and 2010 there had been a relatively high increase in the number of new microbrewery openings, corresponding to a total number of 255 nationwide (see Figure 3).

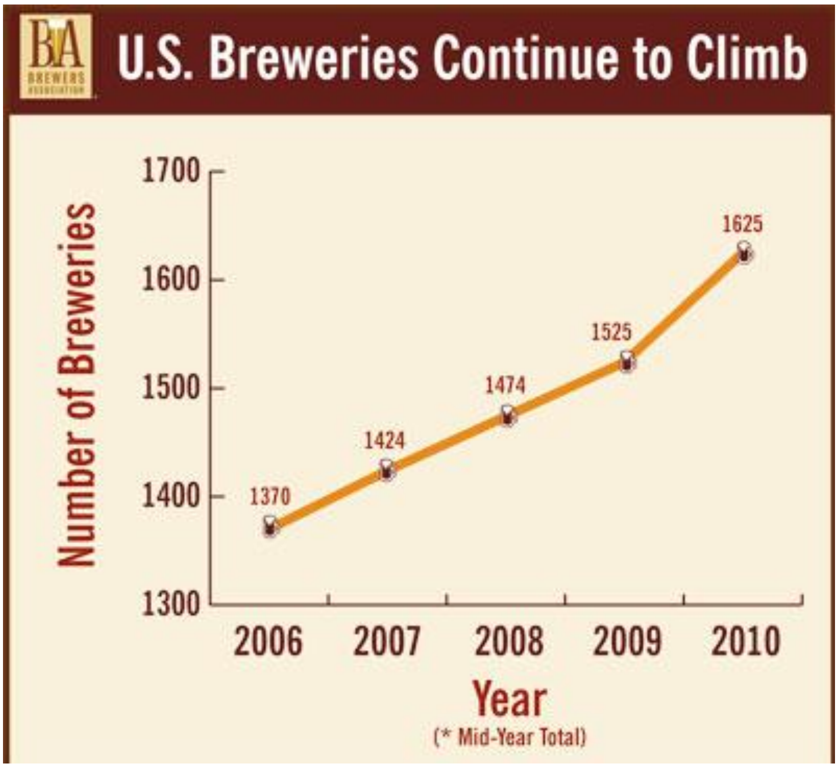


Figure 3: *Number of existing microbreweries in the US from 2006-2010 (Brewers Association).*

The beer products produced by microbreweries are referred to as craft beer, which differs from mass-production beer (i.e. Coors, Miller or Budweiser) in quality, flavor and price. The prices of craft beer are usually set higher than the ones of mass-production beer. This leads to the assumption or hypothesis that typical craft beer consumers originate from higher income groups which in turn also correlates to a higher age of consumers. Mass-producers are unable to efficiently distribute a large variety of beers, which allows microbreweries to target consumers looking for new malt beverages, and unique tastes (Carroll and Swaminatha 2000). The growing number of microbreweries dwarfs the number of mass-producers, which resembles a true change in consumer preferences and a shift in demand for premium craft beer (Parcell and Woolverton 2008).

Beer consumers have started to value the high quality of craft beer which goes along with a gradual increase in craft beer consumption over the past years and this trend is expected to hold up in the future due to an improvement of the economy, innovation of products and an engaged consumer base. According to the market research company Mintel, a year-to-year growth trend is predicted for the next years, with craft beer sales approximated to $36 billion in 2019, which would be an increase in sales of $28 billions from 2012.

1.2.2 Social Media Network Twitter

A major core of this research is the analysis of data pulled from the social media network Twitter to harvest new knowledge of the spatial and demographic characteristics of craft beer consumers.

Twitter was created in March 2006 by Jack Dorsey, Evan Williams, Biz Stone and Noah Glass and by July 2006 the site was launched. The service rapidly gained worldwide popularity, with more than 100 million users who in 2012 posted 340 million tweets per day (Twitter 2012). Twitter is a free social networking micro blogging service that allows registered members to broadcast short messages called *tweets* (TechTarget 2014). These short messages are public by default and tweets can be accessed either via live streaming or the historic data depository. Twitter offers an Application Programming Interface (API) where the respective datasets can be retrieved.

Compared to other social media services, Twitter has attracted relatively more research interests due to its two distinct platform characteristics: tweets can be posted using various tools and the majority of messages are open to the public (Chen et al 2013). The additional availability of locational information associated with the content of social media data widens the analysis scope to the geospatial domain, referring to the analysis of where people disclose the content. Tweets containing a geographical footprint are referred to as geo-referenced tweets. This data corresponds to the x- and y-coordinate where the tweet is posted. The users can decide whether they want to include information about their current location and only about 3% of all the tweets released around the globe are geo-referenced.

A tweet usually contains information about the time when it was released, the author (user), textual content of the author (user), and (if geo-referenced), information about the quantitative (coordinates) and qualitative location (place names) where the tweet is posted. This enables the conduction of a wide range of analysis such as the spatial distribution of tweets regarding a certain topic, applying text-mining techniques on the textual content, study of user characteristics or exploring spatiotemporal patterns on how people move through space over a certain period of time. The output of the analysis can then be used for finding new locations for microbreweries, either for the development of automated generic models or analytical on-the-fly site selection approaches. For this research the former approach is chosen to select new feasible locations.

1.2.3 Geodemographics

Geodemographics establish a relationship between a specific geographical area and the characteristics of the people that live within that area. As a definition, the term can be split in two parts: ***'Geo'*** referring to the spatial characteristics, such as the geographical delineation of a certain group of people and ***'demographics'***, alluding to the attributive characteristics of that group of people. These attributive characteristics most often correspond to census data about the population such as gender, age or race. But demographics can also be non-census based such as data about income, unemployment, or spending power which is referred to as socio-economic data. There's another non-census based data group called lifestyle or behavioral data which includes data about consumer buying habits, product preferences or attitudes toward a certain behavior such as craft beer consumption. The data is usually not available at the individual level (single person) but as an aggregated dataset classified by geographical areas at different scales (ZIP codes, tracts, blocks, etc.) according to the respective demographic characteristics.

(Harris, 2003) defines the nature of geodemographics as the analysis of socio-economic and behavioral data about people, to investigate the geographical patterns that structure and are structured by the forms and function of settlement. Out of this definition, it can be inferred that geodemographics also comprise the study of settlement patterns and structures, their characteristics and how they evolve. An established method both at the commercial and academic sector is called geodemographic segmentation, where settlement structures are described by various clusters distinguished by different criteria such as people's physical characteristics (age, gender, race, etc.), social status or level of wealth. Geographical areas are classified by a conglomerate (cluster) of multiple demographic variables. In the literature, this approach is also often referred to as neighborhood classification. It is based on the principle of Tobler's first law of geography: Everything is related to everything else, but near things are more related than those far apart (Tobler, 1970). Derived from that law, one can assume that people living in the same neighborhood hold similar characteristics. This assumption might be valid on a broad scale, but needs to be taken with caution due to its generality and suppression of population diversity within a neighborhood. (Harris et al 2005) refer to this as the ecological fallacy, which can be understood as making an inappropriate assumption that any specific individual necessarily shares the general characteristics of her or his neighborhood and its population.

Geodemographic segmentation is widely used in areas like marketing, sales forecast, site suitability analysis and is often incorporated in site selection processes. It has come into use as a shorthand label for both the development and the application of area typologies (neighborhood classifications) that have proved to be powerful discriminators of consumer behaviors and aids to market analysis (Brown, 1991). Various commercial software tools have been developed in the past with Tapestry (ESRI), Prizm (Claritas) and Mosaic (Experian) being the most famous ones. Most of these commercial products use a k-means clustering algorithm to create a specific number of groups (typologies) and allocate each neighborhood to the respective group. This allocation is based on the principle that if two neighborhoods share similar demographic characteristics then they'll be assigned to the same group.

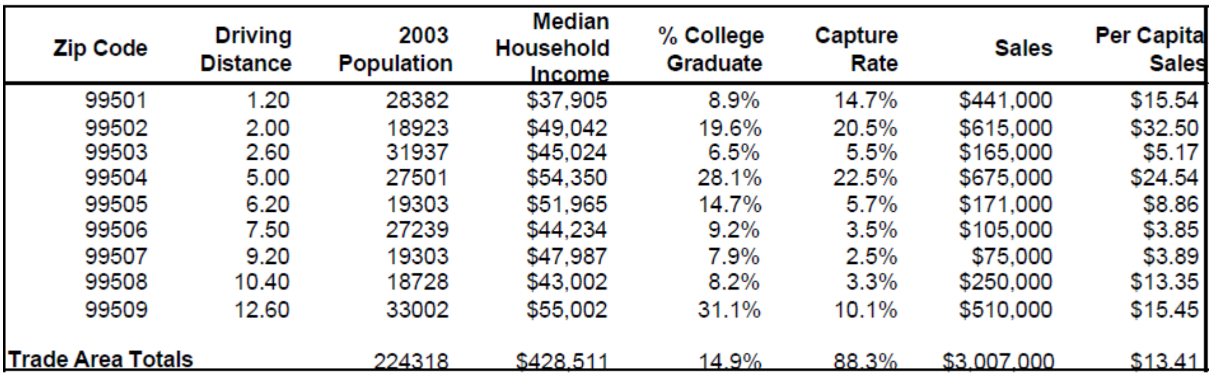
Although k-means is the most commonly established clustering algorithm in geodemographic segmentation, there are multiple other clustering techniques that can be applied such as artificial neural networks, genetic algorithms or fuzzy logic, which are more efficient methods within large, multidimensional databases (Brimicombe 2007). For this research project, a novel geodemographic approach is applied which is less generic and is more resistant to population diversity. This will be described in section 1.5 in more detail.

1.2.4 Traditional Site Selection Models

Numerous site selection models have evolved in the past which serve as an instrument in the process of finding new locations for businesses. This section summarizes the most popular and frequently used site selection models in the retail and restaurant industry.

Analog Model

The analog model focuses on the study of existing retail stores to identify potential retail sites. It constituted the first attempt at a formal retail site selection process (Applebaum 1968). Consumers of existing stores or restaurants are interviewed to obtain knowledge about their characteristics, most often regarding their residential location or their demographic attributes. Determining where customers live enables the delineation of primary catchment areas. The model itself is similar to a regression model. A number of exploratory variables (e.g. Median Household Income, Educational Attainment etc.) are used to estimate the sales potential of a particular site. To determine a likely performance of a planned retail store or restaurant, a systematic comparison of the characteristics of the proposed store with the characteristics of the existing 'analogue' store is performed (Breheny 1988). In practice, the variables and the respective sales level are usually summarized in analog tables and grouped by geographical areas (e.g. ZIP codes). Figure 4 shows an example of an analog table.



Zip Code	Driving Distance	2003 Population	Median Household Income	% College Graduate	Capture Rate	Sales	Per Capita Sales
99501	1.20	28382	$37,905	8.9%	14.7%	$441,000	$15.54
99502	2.00	18923	$49,042	19.6%	20.5%	$615,000	$32.50
99503	2.60	31937	$45,024	6.5%	5.5%	$165,000	$5.17
99504	5.00	27501	$54,350	28.1%	22.5%	$675,000	$24.54
99505	6.20	19303	$51,965	14.7%	5.7%	$171,000	$8.86
99506	7.50	27239	$44,234	9.2%	3.5%	$105,000	$3.85
99507	9.20	19303	$47,987	7.9%	2.5%	$75,000	$3.89
99508	10.40	18728	$43,002	8.2%	3.3%	$250,000	$13.35
99509	12.60	33002	$55,002	31.1%	10.1%	$510,000	$15.45
Trade Area Totals		224318	$428,511	14.9%	88.3%	$3,007,000	$13.41

Figure 4: *Example of an analog data table showing sales potential of each ZIP code and the respective exploratory variables on the ZIP code level (Source:Brubaker 2001).*

The analog model fits well for site selection scenarios where existing data is already available or can be obtained easily and accurately. The advantage of this model is its high adaptability to virtually all types of retail and restaurant establishments. One weakness of this approach is the high individuality which means that its application requires experienced analysts. Another drawback is the fact that it most often requires the development and maintenance of a comprehensive database.

Gravity Model (Huff Model)

Gravity models are often used in site selection for estimating consumer flows, number of potential customers and in more advanced versions for forecasting sales. More abstractly, it describes the interaction between two areas, the origin (e.g. location where the customer lives) and the target location (e.g. store, restaurant) based on the population size of each location and the distance between those. Gravity models fall into the category of spatial interaction models, which have been used by (Sen et al 1995) and others to label models that focus on flows between origins and destinations. (Meyer 1988) defines the gravity model as a method of evaluating human behavior that measures the likelihood that individuals will gravitate toward a store depending on the individuals' travel distance to alternative stores, and the inherent drawing power of each location.

A general gravity model can be mathematically defined as follows:

$$I_{ij} = k * \frac{p_i * p_j}{{d_{ij}}^{\gamma}}$$

with the following notations:

i … origin
j … target
I_{ij} … interaction between areas i and j
p_i … population of area i
pj … population of area j
k … proportionality constant
γ … distance decay factor

This formula was developed by (Carey 1858). It basically states that the interaction between the origin and the target location decreases as the distance between the two location increases. The degree by which the interaction decreases is controlled by the distance decay factor α. Moreover, the interaction rises as the population increases.

A more specific spatial interaction modeling approach is the **huff model**, which is widely used in practice. It is a mathematical specification of the probability that a customer will shop at competing stores or retail centers based on a derivative of the gravity model (Church 2009). Given a network of stores and a set of customer locations, the model calculates the probabilities of each customer patronizing each store. These probabilities can be used as a basis for estimating sales potential, delineating trade areas or generating market areas. There are two versions of the Huff model: the **original huff model** and the **advanced huff model**. The original huff model is based on a mathematical function which approximates the probability that a particular customer will shop at a specific store. The function takes as an input the distance between the residential locations of the customers to the store, the attractiveness of the site and the distance and attractiveness of competitive sites. Mathematically the model can be expressed with the following formula:

$$\alpha_{ij} = \frac{\frac{s_j}{d_{ij}^{\gamma}}}{\sum_{k=1}^{m} \frac{s_k}{d_{ik}^{\gamma}}}$$

with the following notations:

α_{ij} … probability that a customer at location i will shop at site j
i … customer location
j … store j
k … store k (competing store)
sj … attractiveness of store j
d_{ij} … distance between i and j
d_{ik} … distance between i and k
γ … distance decay factor

The original huff model was developed and made public in 1964 by Dr. David Huff of the University of Texas. He suggested taking the size of the store as a measure of attractiveness (Huff 1964). This is based on the assumption that the bigger the store, the larger the selection and the greater the attraction to customers. But this assumption has to be applied with care and is not applicable for all types of establishments. It is applicable for the most types of retail stores but is rather not feasible for restaurants or microbreweries.

An alternative is the advanced huff model which was developed by ESRI. It differs from the original huff model in that it incorporates more than one measure of attractiveness for a site. The mathematical formula of the advanced huff model looks as follows:

$$\alpha_{ij} = \frac{\dfrac{\prod_{h=1}^{H} s_{hj}}{d_{ij}^{\gamma}}}{\sum_{k=1}^{m} \dfrac{\prod_{h=1}^{H} s_{hk}}{d_{ik}^{\gamma}}}$$

The terms s_{hj} and s_{hk} denote the h^{th} attractiveness characteristic of the stores j and k. All huff model inputs, exponents, trade area size, and results require detailed analysis by someone who is well versed in the operation of such a model (ESRI 2014).

Advantages of the huff model include its simplicity in application due to the few data points required to run the model. (Nelson 1958) notes that gravity models provide the advantage of working well with typical, simple situations that use conservative calculations, compared to other models. A major drawback of the huff model is that it is frequently inaccurate due to the fact that only the two parameters store attractiveness and distance between consumers and stores are taken into account to determine consumer habits. Another disadvantage is that the model does not take into consideration the type of consumer transportation especially in respect of walking and public transportation.

Multiple Regression Model

The multiple regression model has evolved from the analog model and holds a similar function. It depicts the relationship between sales potential (**dependent variable or response variable**) and a number of characteristics (**independent variables**) such as store characteristics, driving time, median income in trade area etc. Once an equation has been produced, it then can be used to forecast the sales turnover for a proposed store by substituting values for the independent variables (Breheny 1988). The general equation can be defined as follows:

$$Y = \underbrace{a}_{\text{Constant}} + \underbrace{bX1 + bX2 \dots bXn}_{\text{Store Characteristics}} + \underbrace{bXn + 1 \dots + bXm}_{\text{Catchment Area Characteristics}}$$

Y = dependent variable (sales potential, revenue, etc.)
X = independent variable (store size, median income, educational attainment, etc.)

Advantages of using regression models include their simplicity in implementation and that they allow more complex relationships to be investigated, are more flexible, and can test numerous variables relatively quickly (Bruckner 1998). Drawbacks of regression models include their high degree of individuality often requiring experienced analysts and the availability of comprehensive data of existing stores.

1.3 Objectives and Expected Results

This research consists of two major parts. First the characteristics and behaviors of craft beer consumers in San Diego County should be examined extensively in order to gain new knowledge about the craft beer clientele. The second part focuses on the development of a site selection model utilizing parts of the knowledge obtained from the exploratory analysis to evaluate candidate sites regarding a new microbrewery opening scenario.

1.3.1 Exploratory Analysis

This will be accomplished by conducting an exploratory analysis including spatial, qualitative and quantitative data analysis methods. The data used for this analysis is retrieved from the social media networking service Twitter, where beer-related tweets were filtered out over a one year period of time within the bounding box of San Diego County. A detailed description of Twitter and the respective data used for this analysis are covered in chapters 2.1 and 3.2 respectively.

In the exploratory analysis the behavioral and lifestyle patterns of people towards craft beer will be examined using data they voluntarily disclose. The goals are to link this data to existing microbreweries, to study the distribution of this data through space and to harvest information of the demographic characteristics of a typical craft beer consumer.

This analysis will be divided into three main parts:

Geospatial Analysis

Each tweet contains a location tag (x-coordinate and y-coordinate), referring to the position where the user triggered the message, which enables the conduction of geospatial analysis. The data will be displayed on a map and a heat-map is generated in order to examine the spatial distribution of people consuming craft beer. Additionally, a point density map is generated displaying the number of tweets per neighborhood and ZIP code. Another task will be to allocate each tweet to an existing microbrewery. The result should be a point map displaying each microbrewery as a graduated symbol with the symbol size relative of the number of tweets per brewery.

Qualitative Analysis

The qualitative analysis includes the examination of the unstructured textual information people disclose in their tweets which will be accomplished by applying text-mining techniques. One goal of this type of analysis is to filter out tweets which are beer-related but do not necessarily relate to craft beer related tweets. In a second part, the remaining craft beer related tweets should be categorized using an ontology-driven approach. The tweets should be categorized inter alia based on the microbrewery name which consumers are referring to in the text, the type of beer, style of beer, etc. This ontology should on the one hand facilitate the process of allocating tweets to a respective brewery and on the other hand serve as a prototype for future work such as for an automated system processing craft beer related Twitter data.

Quantitative Analysis

The goal of the quantitative analysis is to gain knowledge about the demographic characteristics of craft beer consumers. The demographic variables examined therefore are gender, age, race, income and religion. Each variable will be divided in classes and the result will be a relative frequency distribution for each demographic variable.

1.3.2 Site Selection Model

The second part of this master thesis project comprises the development of a site selection model for potential new microbrewery opening scenarios. The model should be designed and implemented in such way that it is applicable for both macro- and micro level scenarios. In a macro level analysis, the feasibility of each ZIP code area for a new microbrewery opening should be examined. This should give planners and decision makers an overview, which areas would be most appropriate to site a brewery and which are unsuitable. The result should be displayed as a choropleth map showing the number of potential customers for each ZIP code and each area should be highlighted as graduated colors respective to its feasibility.

The micro level analysis is based on a larger scale where a suitability analysis should be conducted for a number of candidate sites corresponding to vacant properties being for sale or rent. The first objective of this site suitability analysis is to retrieve a score for each candidate site corresponding to the number of potential local customers within the catchment area of each site. For the macro level analysis (ZIP code level), the model should take the geodemographic profile of the ZIP code area as an input, consisting of three demographic variables (gender, age, and race) and output the number of potential customers within that area as a final score. The model logic for approximating the number of potential customers should also be equivalent for the site level. It should correspond to a probability model utilizing the information harvested from the preceding exploratory analysis on the demographics characteristics of craft beer consumers.

A catchment area should be generated around every candidate site which is the area where the vast majority of expected customers originate from. In the next step of the model, the demographic profile of the catchment area of the candidate site will be inputted and the output should be the number of potential customers within the local catchment area, which is equivalent to the ZIP code level analysis.

1.4 Research Scope/Questions

1.4.1 Exploratory Analysis

The scope of the first part of this research project comprises the conduction of a descriptive analysis of beer- and specifically craft beer related consumer data retrieved from the social media network Twitter. Part 1 of this research project includes the following tasks and research questions:

Research Tasks:

1. Examine spatial distribution of consumer data

2. Analyze the point density of tweets per neighborhood to get an understanding which areas indicate a high or low demand on craft beer products respectively

3. Deploy text-mining techniques to analyze the textual content of the consumer data

4. Develop a topic-relevant ontology based on keywords appearing in the textual content to
   a) filter out tweets which do not have relation to craft beer
   b) allocate tweets to existing breweries
   c) serve as a groundwork for potential future systems analyzing craft beer related consumer data

5. Harvest new knowledge about the demographic characteristics of the craft beer clientele

Research Questions:

1. How should the distribution and the densities of consumer data be represented in a meaningful way? (Research Tasks 1. and 2.)

2. Which kind of text-mining techniques should be deployed to fulfill the research task? (Research Tasks 3. and 4.)

3. How can demographic information of craft beer consumers be pulled out of the available consumer dataset and in which way should it be represented in order to be meaningful? (Research Task 5.)

1.4.2 Site Selection Model

The research scope of the second part of this master thesis project covers the design and implementation of a site selection model for new microbrewery opening intentions. Knowledge gained from part 1 about the demographic characteristics of craft beer consumers should serve as a pillar in the development process of the site selection model. Part 2 of this research project includes the following tasks and research questions:

Research Tasks:

1. Assess average catchment area for a microbrewery

2. Develop a geodemographic model for all scale levels (ZIP Code, Blocks, catchment Area, etc.) which takes as an input the geodemographic characteristics of the geographical area and outputs the number of potential customers in that particular area.

3. Utilize the geodemographic model to estimate the number and proportion of potential customers for each ZIP code of San Diego County

4. Extent the geodemographic model developed in task 2 to estimate the potential customers for a number of vacant candidate sites in San Diego County.

Research Questions:

1. How are the local customers of a microbrewery distributed through space and how can the respective catchment area be delineated? (Research Task 1.)

2. How can the number of local potential customers be approximated based on information of craft beer consumption behaviors pulled from external statistical resources combined with the knowledge gained from the exploratory analysis of part 1 of this research project? What modeling approach can and should be chosen to accomplish this? (Research Task 2.)

3. How can the modeling approach of research question 2 be extended in order to be applicable for both macro level site selection scenarios (number of potential customers in ZIP code) and micro level suitability analysis (number of potential customers per vacant candidate site)? (Research Task 3. and 4.)

1.5 Methodology

For the first core part of this research, the exploratory analysis, a heat map will be generated as a first step to get an understanding how the beer-related tweets are distributed through space. To accomplish this, the Kernel Density Function is used, which takes the data points as an input and generates a countywide raster density map showing areas from high to low beer consumption tendencies. For the analysis of the textual content of the tweets, a topic related ontology is developed and qualitative text-mining techniques are applied such as Natural Language Processing (NLP), word extraction and word frequency analysis. The analysis of the demographic characteristics of craft beer consumers will be based on a selection of a sample of a 100 Twitter users frequently posting craft beer related tweets. The public Twitter user profiles are linked to the ReferenceUSA lifestyle database and the respective demographic characteristics are retrieved from that platform. These characteristics comprise the demographic variables gender, age, race, income, religion, and marital status.

As the second core part of this research, a geodemographic model is developed based on the demographic characteristics gained from the previously conducted exploratory analysis. For the implementation of the model, a heuristic approach is chosen to approximate the number of potential local customers. This approach is based on the following principle:

number of potential customers in demographic group
=
number of people in demographic group
(model input → census data)
x
probability that a person of that demographic group is a craft beer consumer
(→ derived from exploratory analysis)

A **demographic group** is defined as follows:

A group of people sharing the same joint demographic characteristics (e.g.: White Males Age 25-34)

A geographical area usually comprises a large number of various demographics groups. Applying the above principle to each demographic group and subsequently summing up all output scores leads to the total number of potential craft beer consumers in that particular area. This can be formalized as follows:

Notations:

C_p … potential number of customers in geographical area (e.g. ZIP code area, catchment area etc.)
DG … demographic group
m … number of demographic groups in geographical area
POP_{DG} … number of people in demographic group
$P(DG_{Craft\ Beer\ Consumer})$ … probability that a person of a specific demographic group is a craft beer

consumer

Model Formula:

$$C_p = \sum_{DG=1}^{m} POP_{DG} * P(DG_{Craft\ Beer\ Consumer})$$

This geodemographic probability model can easily be adopted and enhanced for site feasibility analysis at different geographical scales. To approximate the number of potential craft beer consumers per ZIP code area (macro level analysis), the geodemographic model is run for each area taking as an input the demographic census data. For the second scenario, the feasibility analysis of single vacant sites, the model calculates the number of potential local customers within the catchment area of each site. Therefore, the model needs to be enhanced to generate a catchment area around each site and to prepare the demographic data so it aggregates to the catchment area. This is accomplished by utilizing proper spatial analysis tools such as Buffer, Intersect or Dissolve.

This heuristic approach was chosen over traditional site selection models, because these models require detailed knowledge about their input and calibration parameters. Due to the limited research conducted in past regarding site selection for microbreweries and the unavailability of performance relevant data of existing breweries, applying generic traditional models might lead to unfeasible output scores. (Bruckner 1998) states that gravity models are primarily used with supermarkets and drug stores. Analog models and regression models require the availability of comprehensive data of existing microbreweries, which is not available for this research project. For this heuristic approach, raw geodemographic data derived from the census is incorporated in the model, geodemographic segmentation is not considered due to the ecological fallacy issue discussed in chapter 1.2.3 and due to the fact that most commercial geodemographic systems are intransparent regarding their clustering techniques and might not include segments which fit the demographic profile of an ideal craft beer consumer.

1.6 Structure of this thesis

The subsequent **chapter 2 – Literature Review** provides a detailed state-of-the-art analysis of the work that has been done so far regarding the analysis craft beer consumption, site selection techniques of restaurants and bars and geodemographic modeling approaches.

Chapter 3 – Exploratory Analysis covers all tasks related to gaining new knowledge about craft consumption behaviors based on an extended analysis of the respective Twitter data available. This includes spatial analysis of the tweets, a quantitative analysis of the textual content of the tweets and a qualitative analysis to determine the demographic characteristics of craft beer consumers.

In **Chapter 4 – Site Selection Model** a novel heuristic model is designed and implemented for new microbrewery opening scenarios. The model is developed based on information retrieved from the exploratory analysis in chapter 3. The model is highly adaptable to different scales of suitability analysis. A macro level analysis is conducted on the ZIP code level to approximate the number of potential craft beer consumers within each ZIP code in San Diego County. In a micro level analysis, the model is adapted to estimate the number of potential customers for various specific sites which are currently for rent or sale in San Diego County.

Chapter 5 – Model Application Scenarios applies the geodemographic model to two site selection scenarios, a macro level scenario and micro level scenario.

Chapter 6 – Conclusion and Future Work summarizes all aspects of this work and delivers possible future enhancements of this groundwork and outlines what additional data analysis task could conducted and how the model can be extended to estimate sales potential for instance.

2 Literature Review

This chapter summarizes some of the related work that has been done regarding the analysis of Twitter data, geodemographics and site selection methods in the retail and restaurant industry.

2.1 Twitter Data Analysis – Related Work

The rise and popularity of the social media network Twitter has generated a massive amount of public data which has drawn the interest of professionals and scholar to analyze the various types of data included in public tweets in order to gain new knowledge about the characteristics and behaviors of people regarding different topics. (Li et al 2013) use Twitter and Flickr data to explore spatiotemporal patterns of geographic data generated in social media, within the bounding box of the contiguous United States and further infers the characteristics of users by examining the relationships between geographic data densities and socioeconomic characteristics of local residents at the county level using California as a case study. In a point map they visualized the spatial distribution of the georeferenced tweets within the United States and parts of the Los Angeles area (see Figures 5 and 6).

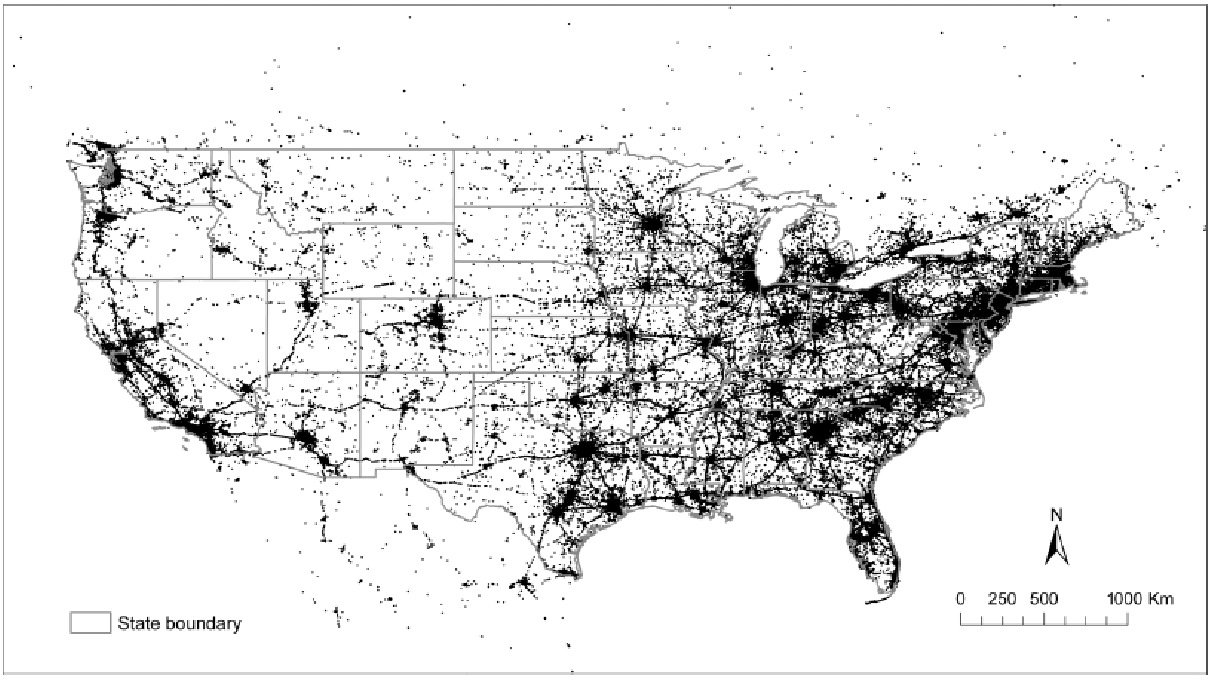


***Figure 5:** Georeferenced tweets within the bounding box of the contiguous United States (Li et al 2013).*

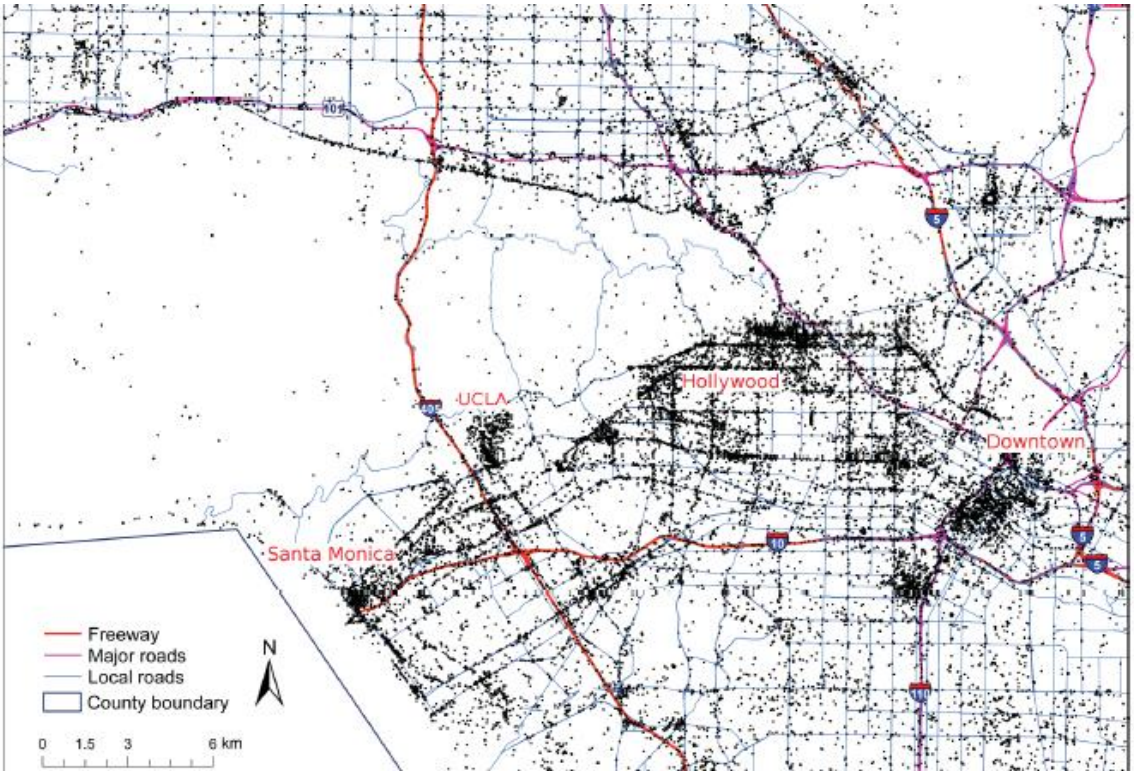


Figure 6: *A close-up of georeferenced tweets in part of Los Angeles (Li et al 2013).*

They applied the kernel density function to calculate and visualize the densities of tweets. A raster-based surface was generated showing the tweet density hot spots (see Figures 7 and 8).

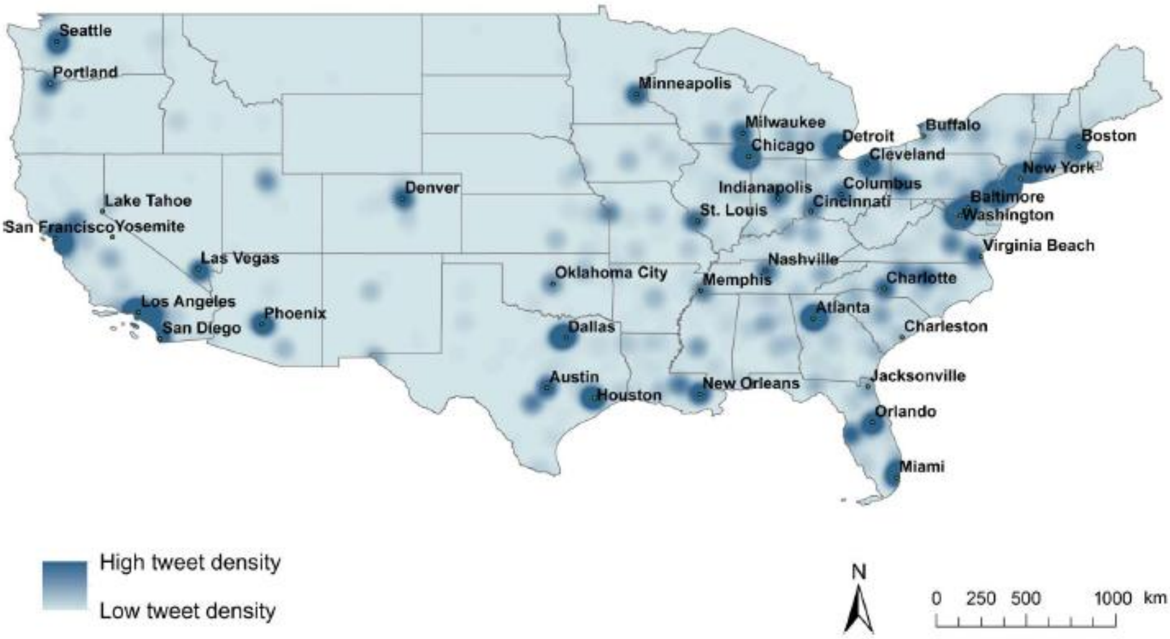


Figure 7: *Tweet density within the bounding box of the contiguous United States (Li et al 2013).*

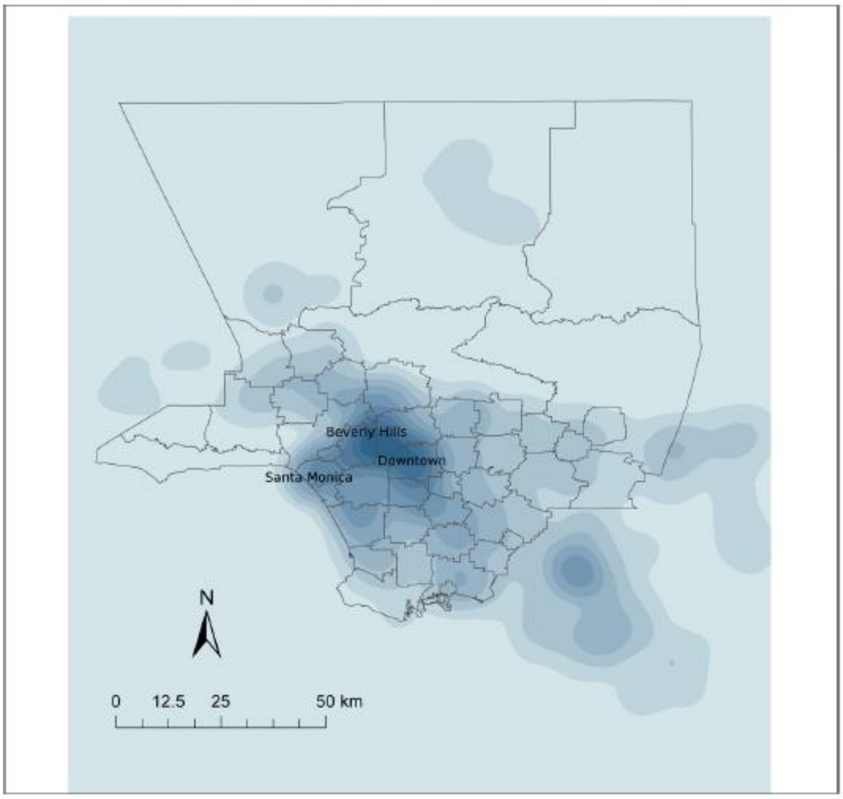


Figure 8: *Tweets density in Los Angeles County (Li et al 2013).*

Additionally (Li et al 2013) examined the temporal patterns of tweets on an hourly basis (see Figure 9).

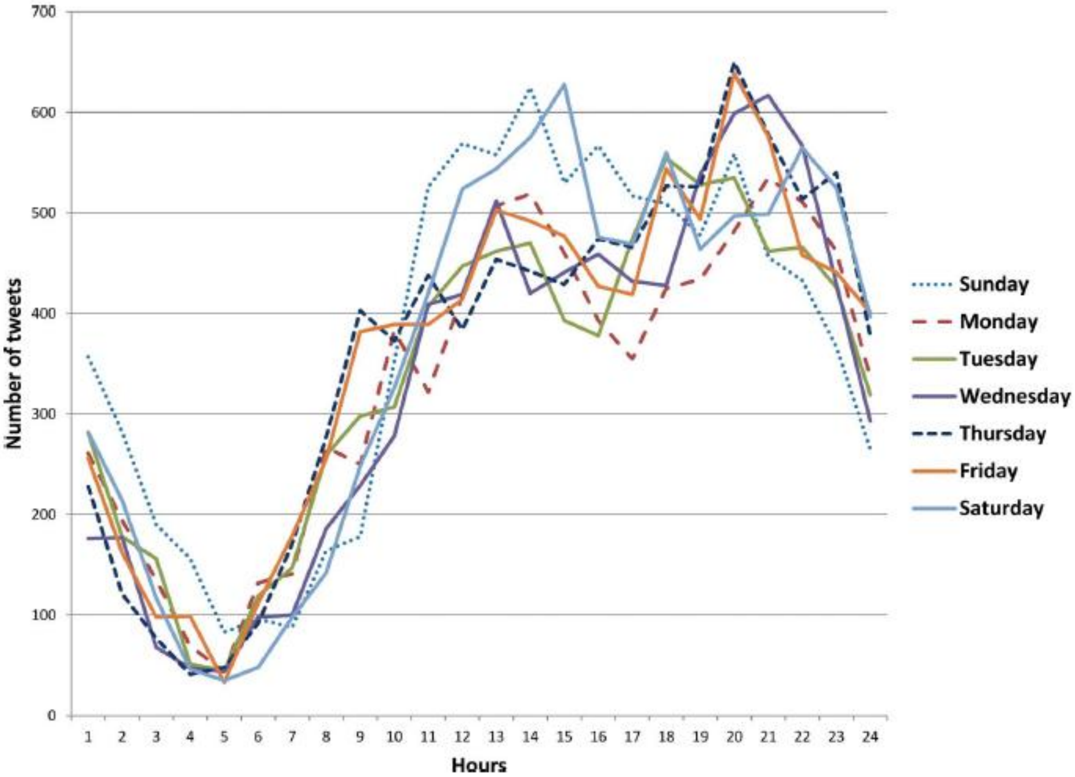


Figure 9: *The average number of tweets per hour in Los Angeles County (Li et al 2013).*

(Fuchs et al 2013) present an exploratory study of the potential of geo-referenced Twitter data for extracting knowledge about significant personal places, behaviors and potential interests of people. The study was done analyzing two months' worth of tweets from residents of the greater Seattle area. Tweets are categorized by distinctive topics such as food, work or education using a minimalistic ontology approach based topic relevant key words. Spatio-temporal aggregation and clustering methods were applied with visual inspection to georeferenced Twitter messages to gain an understanding of significant personal places; as well as to find communities of people with similar interests and analyze their movement patterns (Fuchs et al 2013). Figure 10 shows the temporal tweet distribution of the tweets classified by topic:

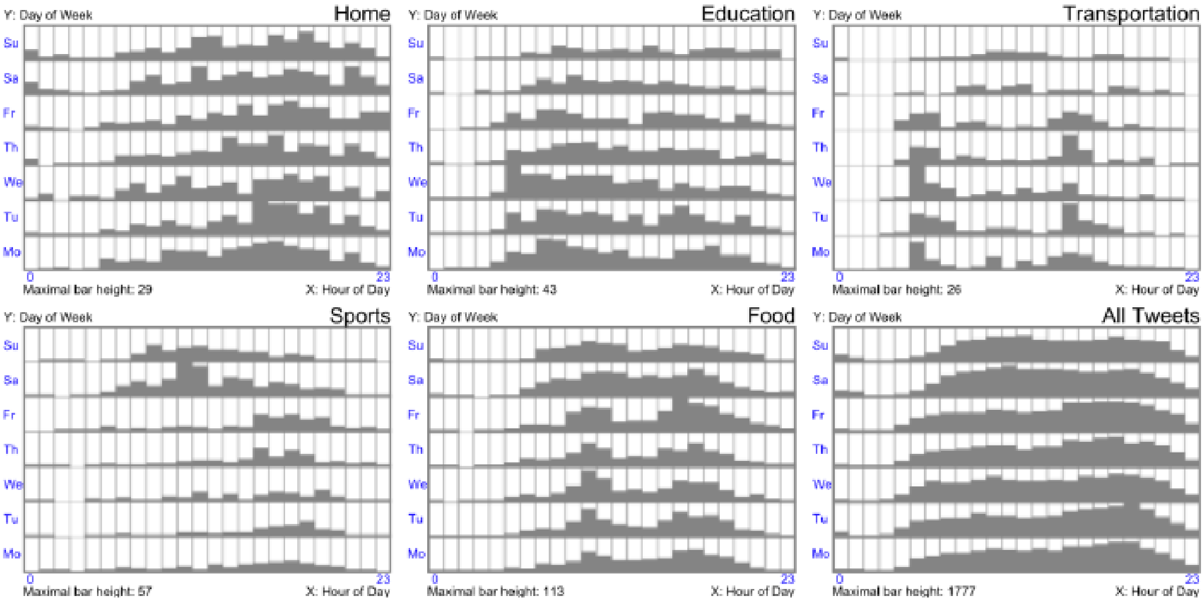


Figure 10: *The temporal distribution of tweets (Fuchs et al 2013).*

K-means clustering was applied to discover clusters (communities) of people with similar interests, represented by combinations of frequently occurring topics (Fuchs et al 2013). Figure 11 shows two examples how different topics are related to each other in respect to how people move from place to place.

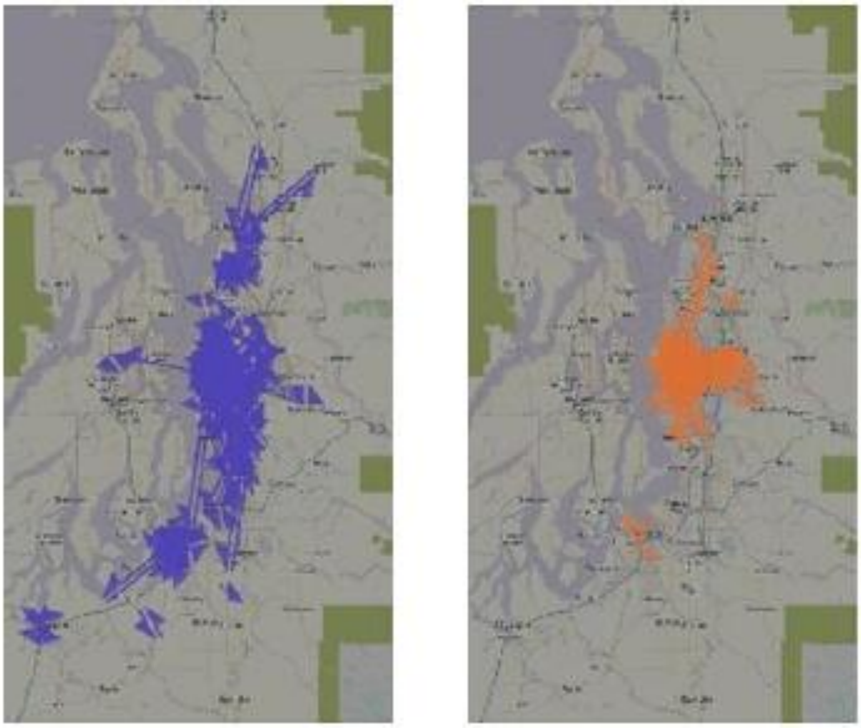


Figure 11: *Clustering of persons' trajectories with at least 10%-dominating topics. Left: "work" with correlated tendency to "food"; right: "food" correlating with "coffee" (Fuchs et al 2013).*

(He et al 2013) describe an in-depth case study which applies text mining to analyze unstructured textual content on Facebook and Twitter sites of the largest three pizza chains in the US: Pizza Hut, Domino's pizza and Papa John's Pizza. The results reveal the value of social media competitive analysis and the power of text mining as an effective technique to extract business value from the vast amount of available social media data (He et al 2013). Figure 12 shows the number of tweets posted on the three different Twitter websites on various days in the month of October.

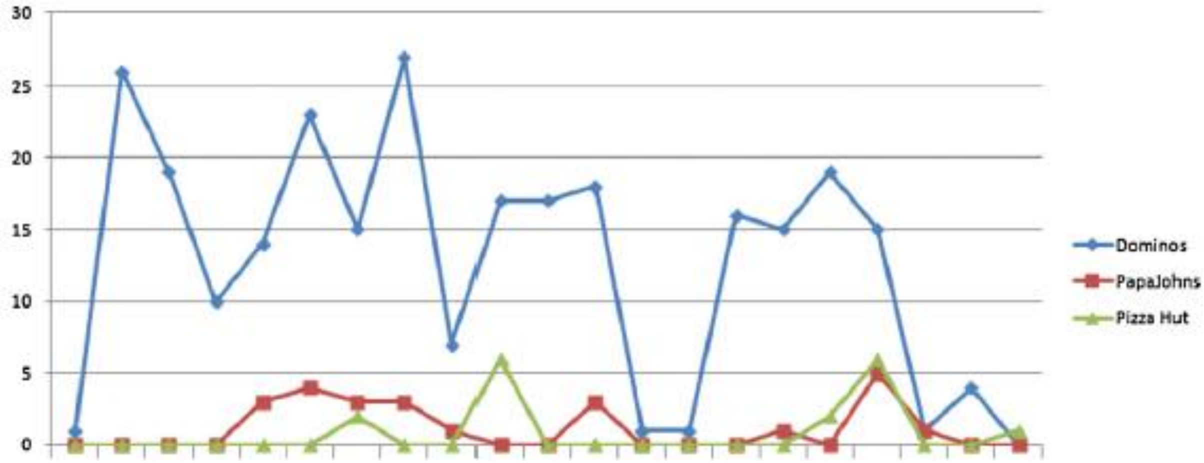


Figure 12: *Trend of tweet numbers in October for the three biggest pizza companies in the US (He et al 2013).*

(Xu et al 2013) conducted an exploratory study of Twitter data examining the geographical awareness in the physical space. They used established text-mining techniques such as NLP to identify place names within the textual content of the tweets. These locations were analyzed in a geographical-hierarchal manner to build a geographical awareness profile for each individual (Xu et al

2013). Additionally they examined the geographical content of tweets and explored if social media data may provide insight about the users geographical knowledge and awareness (Xu et al 2013). They mapped the frequency of place names occurring in tweets and the awareness of place names other than the users' residential place name (see Figure 13).

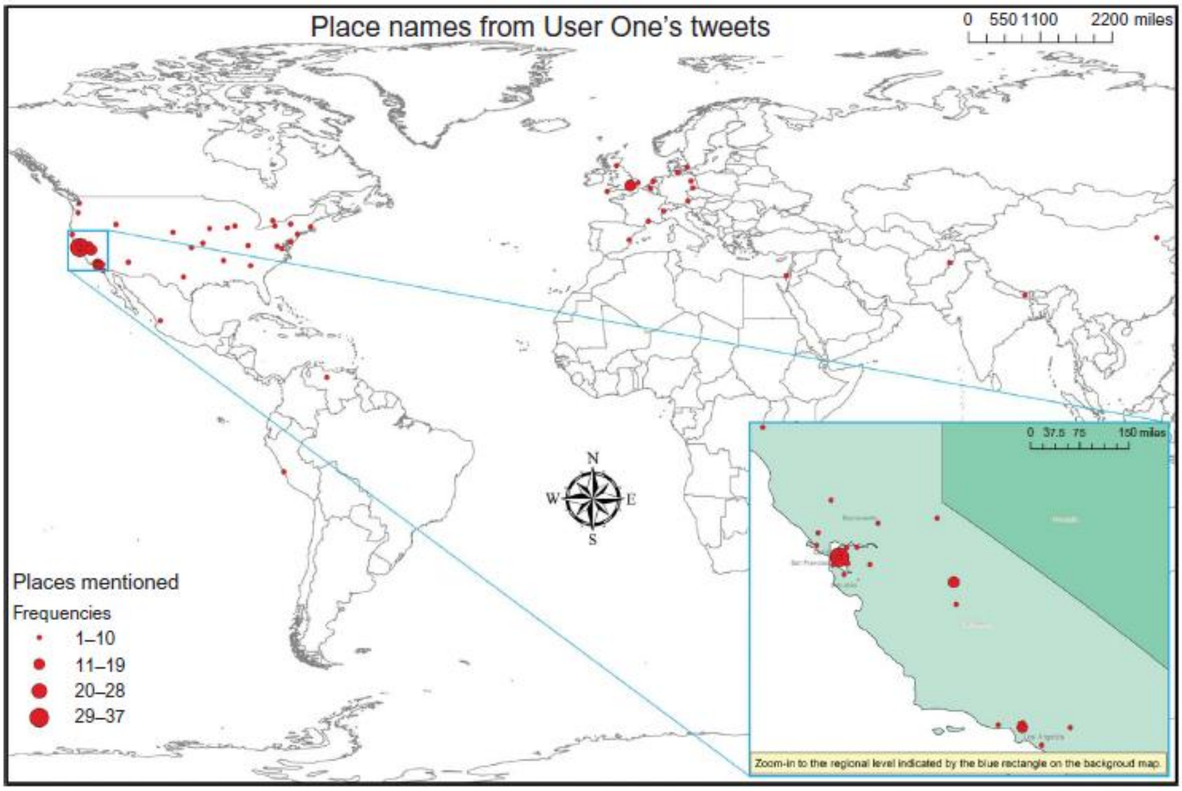

Place names from User One's tweets
0 550 1100 2200 miles
N
W
E
S
Places mentioned
Frequencies
1–10
11–19
20–28
29–37
0 37.5 75 150 miles
Zoom-in to the regional level indicated by the blue rectangle on the backgroud map.

Figure 13: *A sample map showing all georeferenced place names of a single user; Symbol sizes weighted by frequencies (Xu et al 2013).*

Figure 14 and Figure 15 show the geographical awareness of two distinct users:

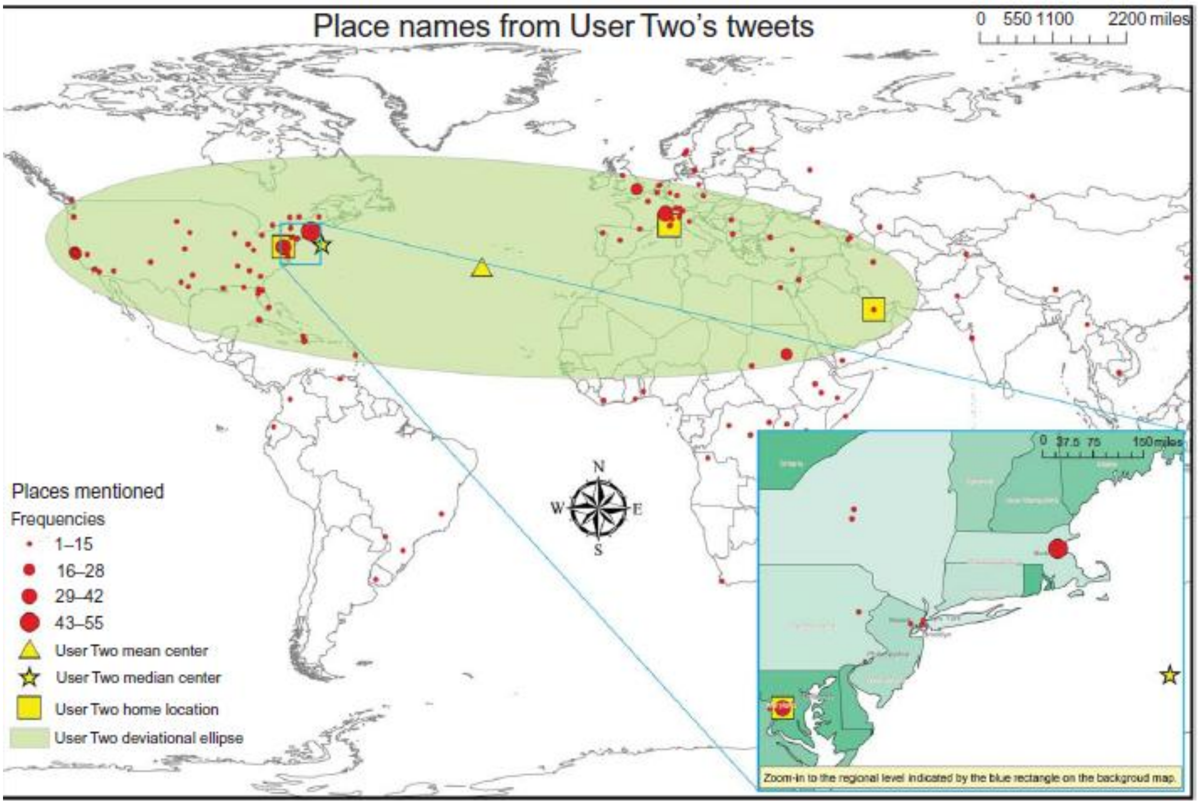


Figure 14: *A user with a very international viewpoint (Xu et al 2013).*

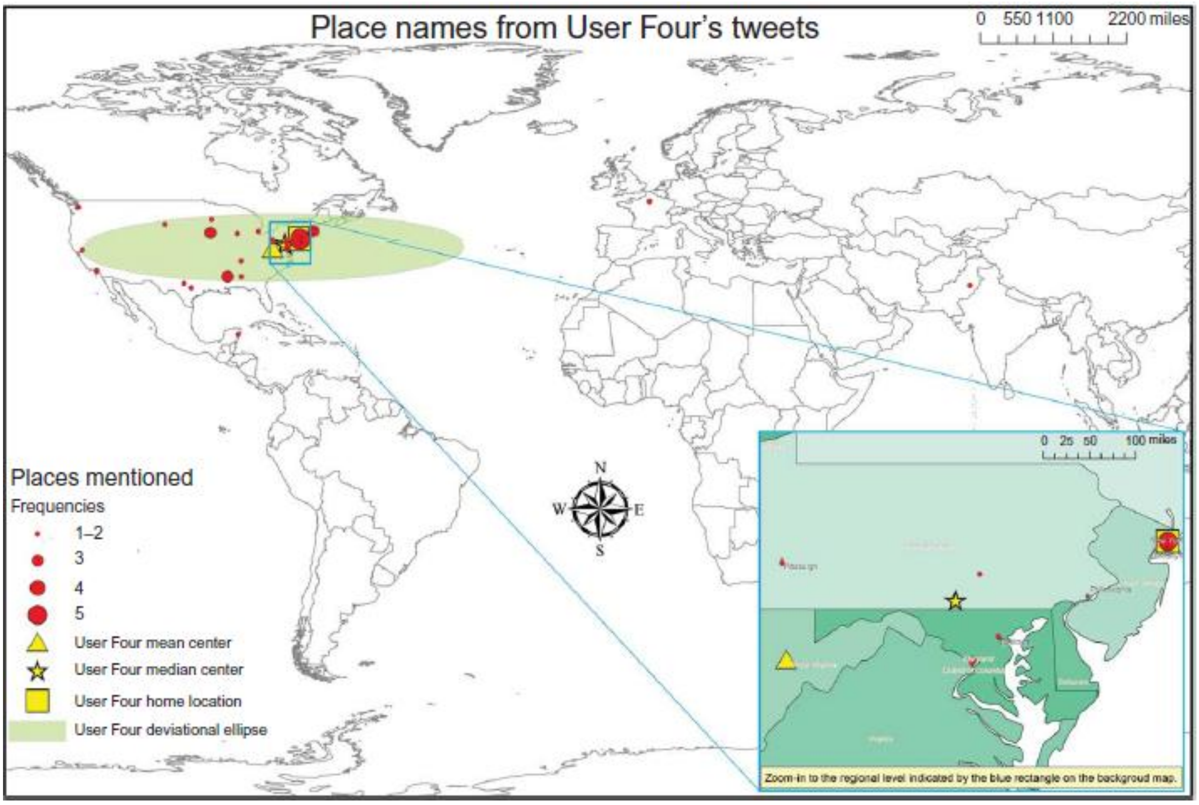

Place names from User Four's tweets
0 550 1100 2200 miles
Places mentioned
Frequencies
1–2
3
4
5
User Four mean center
User Four median center
User Four home location
User Four deviational ellipse
N
W
E
S
0 25 50 100 miles
Zoom-in to the regional level indicated by the blue rectangle on the backgroud map.

Figure 15: *A user with regional and local attentions (Xu et al 2013).*

(Ghosh et al 2013) use data retrieved from the Twitter API to analyze spatial patterns regarding obesity within the United States. The textual content of tweets was analyzed with "R" statistical software packages to identify topics which relate to obesity. Figure 16 shows a point density map which visualizes the spatial distribution of obesity related tweets based on different search terms.

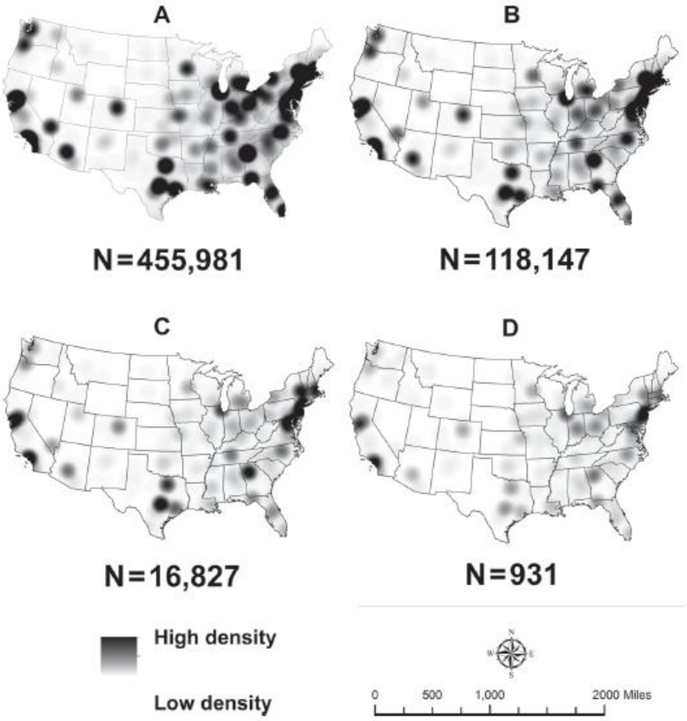


Figure 16: *Visualization of geocoded tweets based on obesity-related search terms. A: All search terms, B: 'Obesity', C: 'Childhood and Obesity', D: 'McDonalds and Obesity'.; N=number of tweets (Ghosh et al 2013).*

2.2 Geodemographics- Related Work

The analysis of the characteristics of people in conjunction with their residential location has attracted a large amount of previous research especially in respect of marketing and site selection. (Lynn et al 1997) provide an overview of the history and development of geodemographics and GIS. GIS development and ways in which textiles and clothing researchers may apply a geodemographic system or GIS in their research are discussed (Lynn et al 1997).

(Adnan et al 2013) undertake a preliminary investigation of the observed differences between the residential geographies of different ethnic groups in the greater London area. They use both geodemographic data obtained from the Electoral Register of Great Britain and data obtained from the Twitter API to examine the distribution of various ethnic groups during both daytime and night time. They analyzed the textual content of tweets and public user profile information to allocate tweets to their respective ethnic profiles. Figure 17 and Figure 18 show the distribution of the various ethnic groups in greater London during day time and night time respectively.

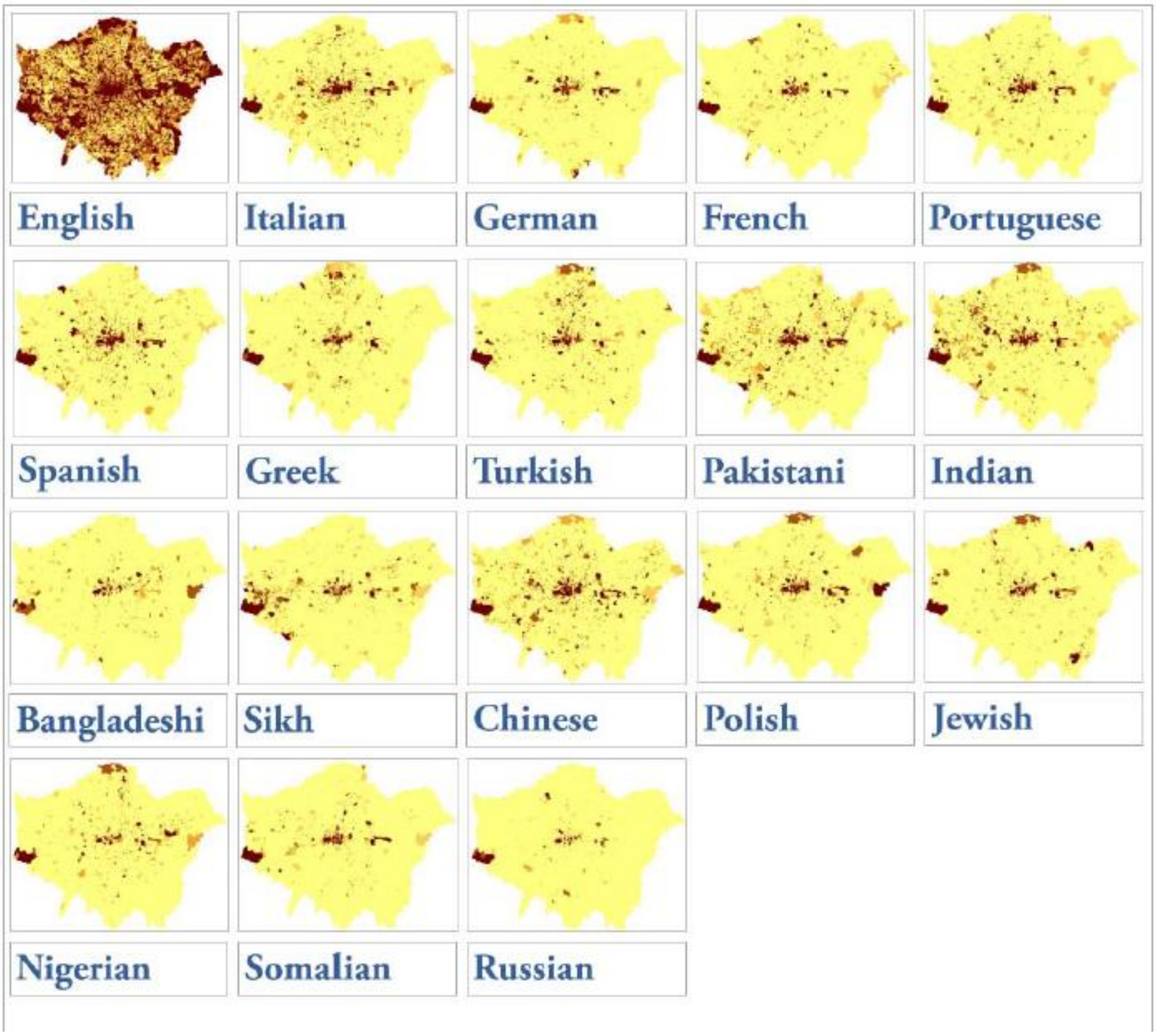


Figure 17: *Distribution of different ethnic groups during day time (Adnan et al 2013).*

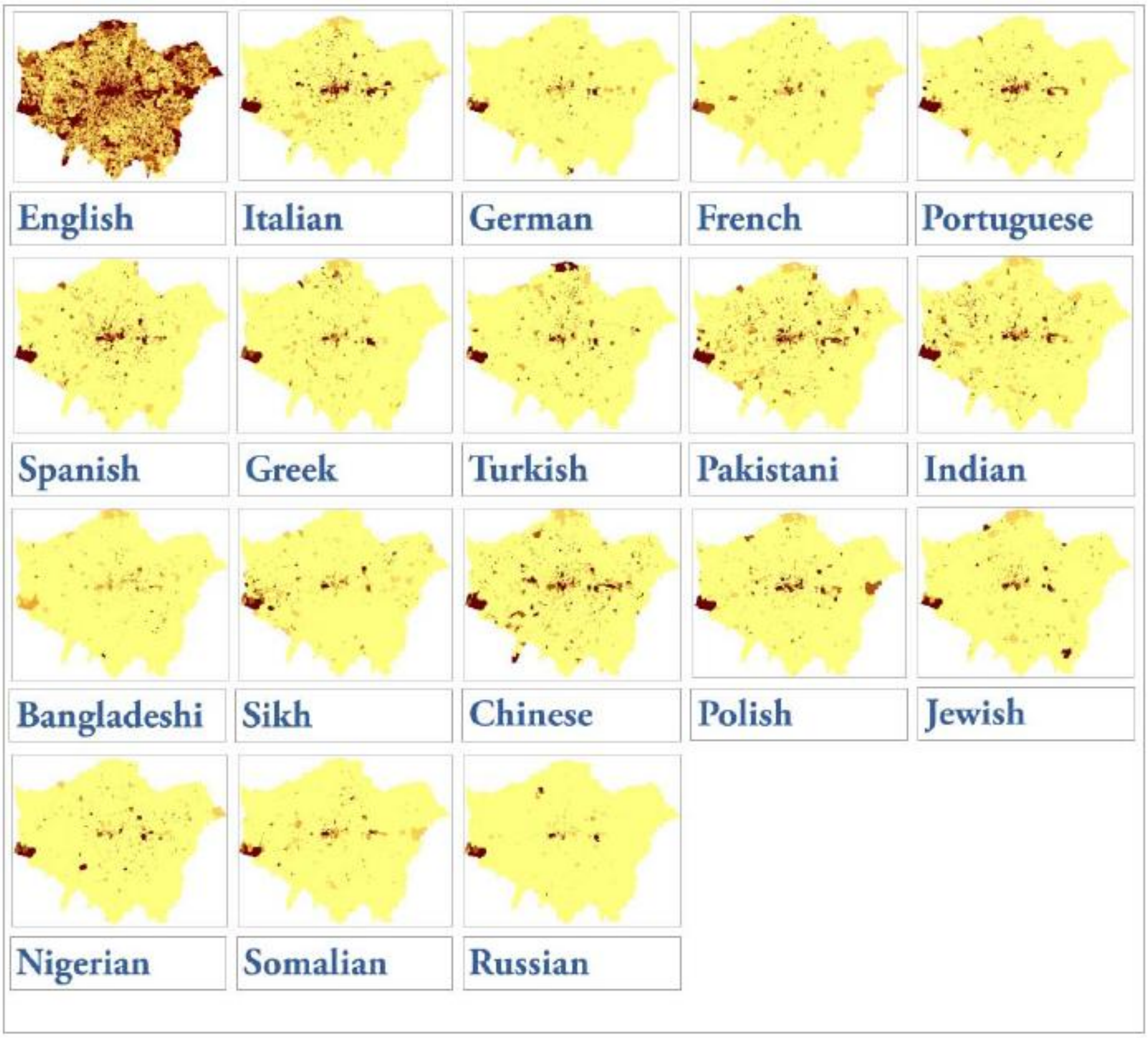
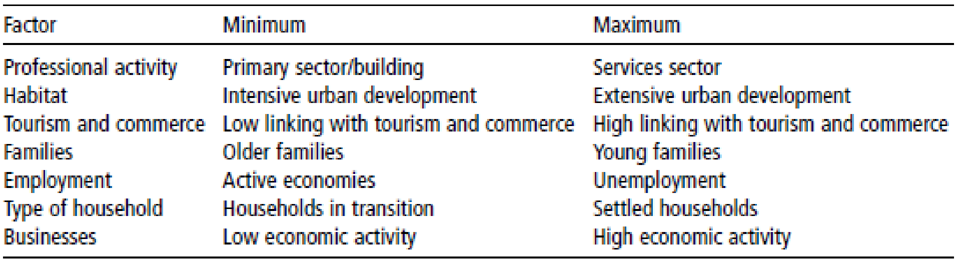


Figure 18: *Distribution of different ethnic groups in the night time (Adnan et al 2013).*

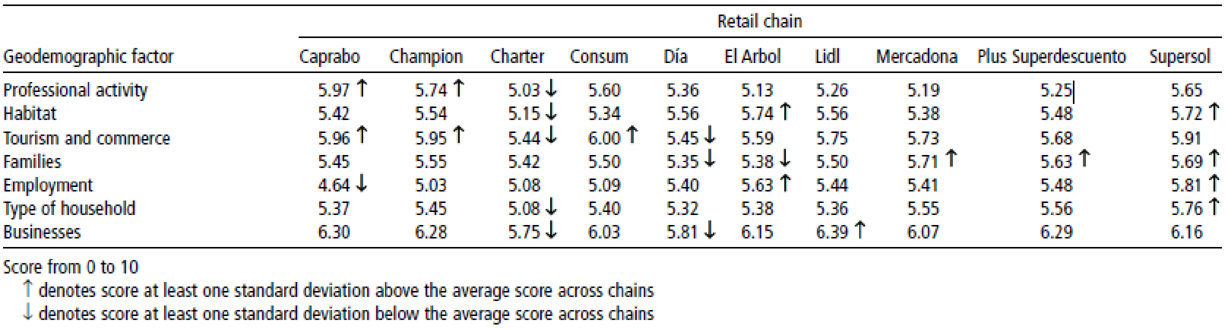
(Gonzalez-Benito et al 2005) study the role of geodemographic segmentation as an analytic tool in retail location strategy. The most relevant factors that should determine retail location selection are revised, and the potential contribution of geodemographic segmentation to the assessment of such factors is examined (Gonzalez-Benito et al 2005). The empirical application provides evidence on the differences between store networks of leading Spanish supermarket chains in relation to the geodemographic profile of their market areas (Gonzalez-Benito et al 2005). The geodemographic segmentation software MOSAIC was used to classify geographical areas based on demographic criteria. MOSAIC classifies neighborhoods based on seven demographic factors (see Table 1).



Factor	Minimum	Maximum
Professional activity	Primary sector/building	Services sector
Habitat	Intensive urban development	Extensive urban development
Tourism and commerce	Low linking with tourism and commerce	High linking with tourism and commerce
Families	Older families	Young families
Employment	Active economies	Unemployment
Type of household	Households in transition	Settled households
Businesses	Low economic activity	High economic activity

Table 1: *Geodemographic classification factors of the segmentation software MOSAIC (Experian Marketing Services).*

The geodemographic profiles of each supermarket chain were analyzed and the respective mean geodemographic factor score calculated for each chain (see Table 2).



	Retail chain									
Geodemographic factor	Caprabo	Champion	Charter	Consum	Día	El Arbol	Lidl	Mercadona	Plus Superdescuento	Supersol
Professional activity	5.97 ↑	5.74 ↑	5.03 ↓	5.60	5.36	5.13	5.26	5.19	5.25	5.65
Habitat	5.42	5.54	5.15 ↓	5.34	5.56	5.74 ↑	5.56	5.38	5.48	5.72 ↑
Tourism and commerce	5.96 ↑	5.95 ↑	5.44 ↓	6.00 ↑	5.45 ↓	5.59	5.75	5.73	5.68	5.91
Families	5.45	5.55	5.42	5.50	5.35 ↓	5.38 ↓	5.50	5.71 ↑	5.63 ↑	5.69 ↑
Employment	4.64 ↓	5.03	5.08	5.09	5.40	5.63 ↑	5.44	5.41	5.48	5.81 ↑
Type of household	5.37	5.45	5.08 ↓	5.40	5.32	5.38	5.36	5.55	5.56	5.76 ↑
Businesses	6.30	6.28	5.75 ↓	6.03	5.81 ↓	6.15	6.39 ↑	6.07	6.29	6.16

Score from 0 to 10
↑ denotes score at least one standard deviation above the average score across chains
↓ denotes score at least one standard deviation below the average score across chains

Table 2: *Characterization of retail chains with geodemographic factors (Gonzalez-Benito 2005).*

(Birkin et al 1998) review the potential of geodemographics to support various activities undertaken by financial institutions and discuss the role that Geographic Information Systems (GIS) can play in enhancing geodemographic products. The aim of this article is to introduce the reader to the nature and use of geodemographics and to demonstrate through examples how spatial analysis can empower both the financial service industry generally and current geodemographic products in particular (Birkin et al 1998). They outline main characteristics of geodemographic products, try to establish a relationship between geodemographics and GIS, and discuss restrictions regarding geodemographics. Existing customers are allocated to their respective geodemographic profiles, which can be very valuable in the process of finding new branch locations. They used geodemographics to predict the flow of people that opened a new bank account at a particular branch office (see Figure 19).

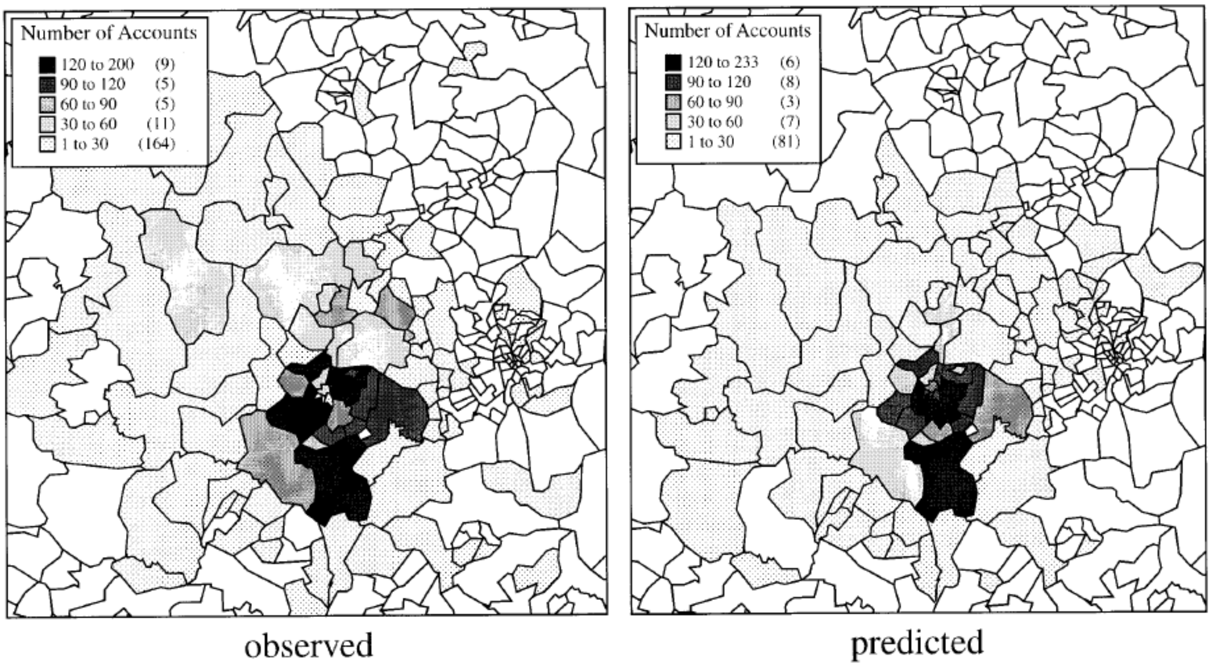


Figure 19: *Flows of savings accounts to a particular bank branch (Birkin et al 1998).*

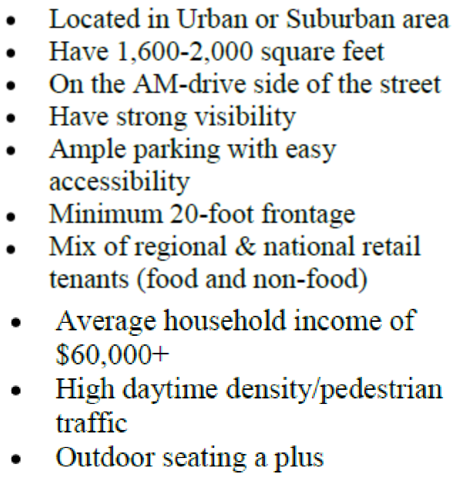
2.3 Site Selection – Related Work

The optimal location of a business is a key factor of its success. Therefore, extensive research both on the academic and commercial sector has been conducted in the past to find feasible methods for location suitability evaluation. (Church et al 2009) integrate traditional location science techniques and modeling with GIS. This book covers both traditional modeling techniques and analytic approaches to assess the feasibility of sites. Major topics include the application of models to analyze existing systems (marketing and distribution), the siting of single facilities (point, line or area), multiple simultaneous siting decisions and models for covering, dispersion/noxious facilities and median/plant location (Church et al 2009). The book also extensively covers the nature and application of gravity models such as the Huff Model, which are discussed in chapter 1.2.4 of this thesis in more detail.

Another pertinent literature is the book of (Thrall 2002) which describes market analysis methods for the real estate market and miscellaneous other business topics which relate to geography such location analysis, the delineations of primary and secondary market area for restaurant businesses, or geodemographics.

(Li et al 2005) introduce in their paper a new approach to site selection that makes use of GIS that incorporates the utilization of the Data Envelopment Analysis (DEA) method called a GIS-based site selection system. DEA is a mathematical programming approach to provide a relative efficiency assessment (called DEA efficient) for a group of decision making units (DMU) with a multiple number of inputs and outputs (Wei 2001). The site selection system described in this paper also focuses on real estate site suitability assessment scenarios.

(Ringo 2009) describes a site selection approach using GIS to determine potential areas where one could establish a Dunn Bros Coffee house franchise. The company suggests in order for a franchise to be profitable, the location must meet the following criteria (Ringo 2009):



- Located in Urban or Suburban area
- Have 1,600-2,000 square feet
- On the AM-drive side of the street
- Have strong visibility
- Ample parking with easy accessibility
- Minimum 20-foot frontage
- Mix of regional & national retail tenants (food and non-food)
- Average household income of $60,000+
- High daytime density/pedestrian traffic
- Outdoor seating a plus

Source: *Dunn Bros Coffee Franchising Inc.*

First the existing branch network is examined based on different demographic criteria. Figure 20 and Figure 21 show some examples.

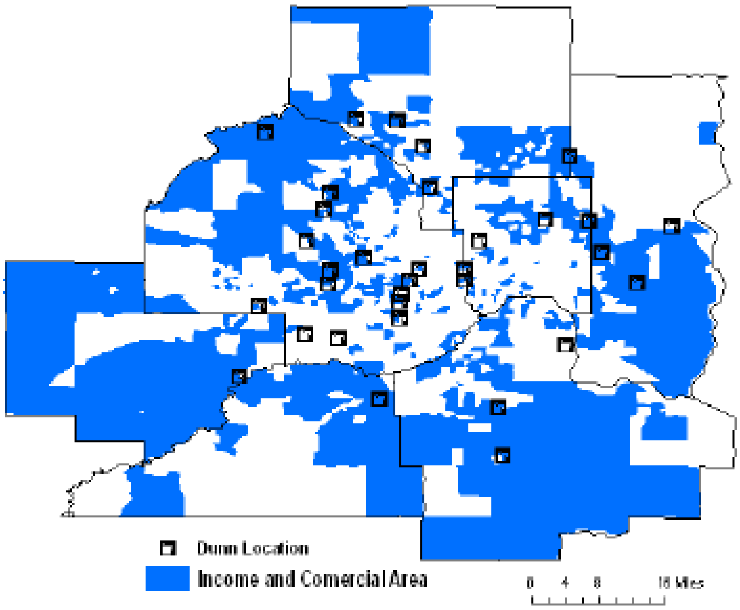


Figure 20: *33 Dunn Bros Coffee store locations that are located within high income (>$60.000) metro block group areas (Ringo 2009).*

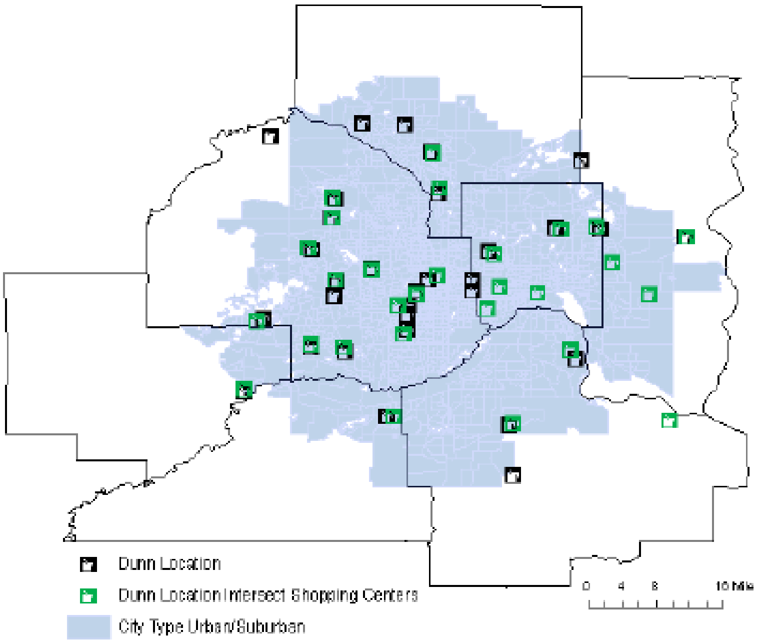


Figure 21: *Dunn Bros Coffee locations within urban areas, shopping centers, and high income block group areas (Ringo 2009).*

The model takes as an input three parameters corresponding to geodemographic data sets (see Figure 22).

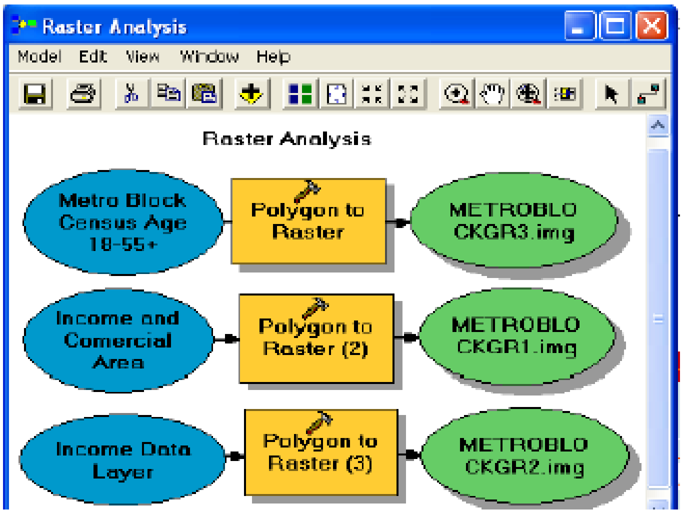


Figure 22: *ArcGIS Model Builder showing a model with three input parameters and three raster suitability layers as an output (Ringo 2009).*

The site selection model corresponds to a pass/fail model which means that there are only two options; either a particular area is suitable or not suitable. The results after running the model are shown in Figure 23.

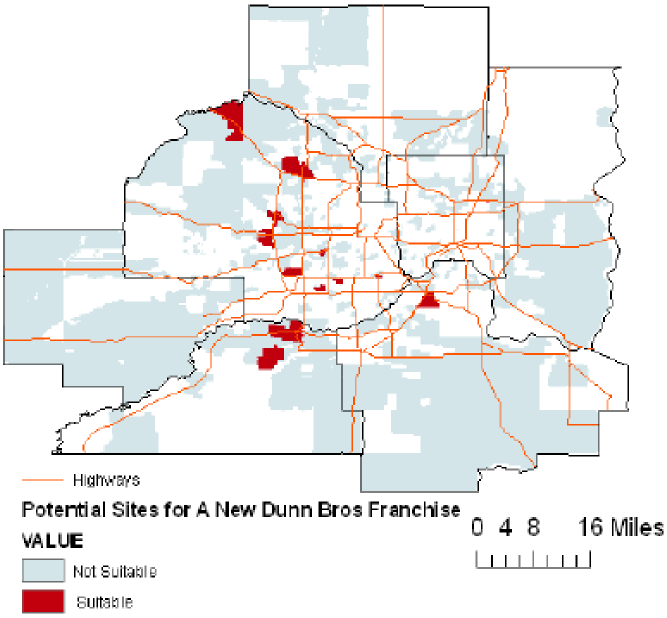


Figure 23: *Suitable locations (red color) to establish a Dunn Bros Coffee franchise. White areas did not fit in the analysis based on the criteria used to conduct this study (Ringo 2009).*

3 Exploratory Analysis

3.1 Overview and Objectives

The rising popularity of social media in combination with the continuous proliferation of mobile devices has led to a massive pool of information generated by people around the world. Thus, people are expressing their desires, feelings and emotions and hence voluntarily sharing parts of their lives. This information can be a valuable source for gaining new knowledge about lifestyles and behaviors of people and hence be powerful in marketing and decision making. As a result many large companies are adopting social media to accommodate this growing trend in order to gain business values such as driving customer traffic, increasing customer loyalty and retention, increasing sales and revenues, improving customer satisfaction, creating brand awareness and building reputation (Culnan 2010).

For this research, data from the social media network Twitter is used to study the behavioral and lifestyle patterns of people regarding the attraction to craft beer products. The analysis conducted on the Twitter data available for this project exclusively utilizes geo-referenced tweets.

Subject of the following sub-sections will be the conduction of an in-depth exploratory analysis of craft beer related and general beer related tweets which had been collected over a period of 10 month. The major goal of this analysis is to retrieve behavioral knowledge about people regarding the demand and consumption of microbrewery products based on the data contained in tweets. The output of this analysis will subsequently be used for the development of the site selection model which will be discussed in chapter 4.

This exploratory analysis is structured as follows:

Section 3.2 – Twitter Data Description gives a brief overview of the structure of Twitter data and the data available for this research

Section 3.3 – Qualitative Analysis describes the analysis of the textual content of tweets using traditional text-mining techniques such as Named Entity Recognition (NER). Additionally a minimalistic ontology is developed based on craft beer related keywords mined from the text.

Section 3.4 – Spatial Analysis comprises the analysis of the spatial distribution of tweets, tweet point densities per neighborhood/ZIP code and number of tweets per existing breweries.

Section 3.5 – Quantitative Analysis undertakes a consumer-specific analysis to determine the demographics characteristics of craft beer consumers such as gender, age or race.

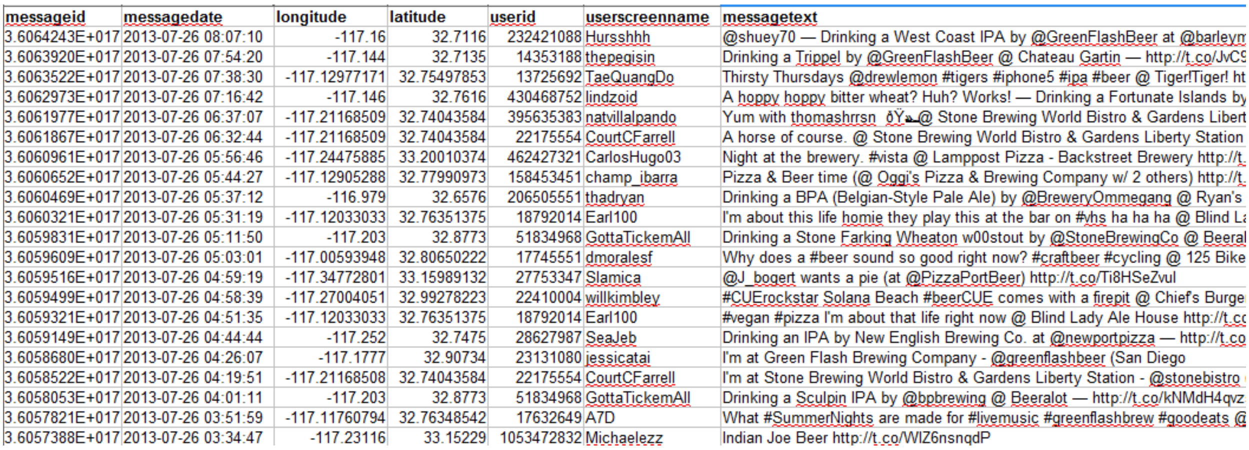
3.2 Twitter Data Description

The data used for this analysis was obtained from the Twitter API. Tweets were fetched over a 10 month period from October 2012 to July 2013 covering the entire area of San Diego County. The data queries were based on topic-related keywords in the textual content of the tweets such as *beer*, *micro brew* or *craft beer.* Only geo-referenced tweets were retrieved from the data depository, topic-related tweets that did not contain locational information were not taken into account for this analysis. Additionally each data entry contains the exact timestamp when the tweet was posted. The data structure of tweets is described in Table 3:

Column Name	Data Type	Description
MESSAGEID	INTEGER	Unique identifier for each tweet
MESSAGEDATE	DATETIME	Date and time when the tweet was posted
LONGITUDE	DOUBLE	Longitude coordinate where tweet was posted
LATITUDE	DOUBLE	Latitude coordinate where tweet was posted
USERID	INTEGER	Unique identifier of the user that released the tweet
USERNAME	STRING	Name of the user
MESSAGETEXT	STRING	Textual content of the tweet

Table 3: The typical data structure of a tweet.

An extract of the beer-related Twitter dataset is shown in Figure 24.



messageid	messagedate	longitude	latitude	userid	userscreenname	messagetext
3.6064243E+017	2013-07-26 08:07:10	-117.16	32.7116	232421088	Hursshhh	@shuey70 — Drinking a West Coast IPA by @GreenFlashBeer at @barleym
3.6063920E+017	2013-07-26 07:54:20	-117.144	32.7135	14353188	thepegisin	Drinking a Trippel by @GreenFlashBeer @ Chateau Gartin — http://t.co/JvC9
3.6063522E+017	2013-07-26 07:38:30	-117.12977171	32.75497853	13725692	TaeQuangDo	Thirsty Thursdays @drewlemon #tigers #iphone5 #ipa #beer @ Tiger!Tiger! ht
3.6062973E+017	2013-07-26 07:16:42	-117.146	32.7616	430468752	lindzoid	A hoppy hoppy bitter wheat? Huh? Works! — Drinking a Fortunate Islands by
3.6061977E+017	2013-07-26 06:37:07	-117.21168509	32.74043584	395635383	natvillalpando	Yum with thomashrrsn ðŸ@ Stone Brewing World Bistro & Gardens Libert
3.6061867E+017	2013-07-26 06:32:44	-117.21168509	32.74043584	22175554	CourtCFarrell	A horse of course. @ Stone Brewing World Bistro & Gardens Liberty Station
3.6060961E+017	2013-07-26 05:56:46	-117.24475885	33.20010374	462427321	CarlosHugo03	Night at the brewery. #vista @ Lamppost Pizza - Backstreet Brewery http://t.
3.6060652E+017	2013-07-26 05:44:27	-117.12905288	32.77990973	158453451	champ_ibarra	Pizza & Beer time (@ Oggi's Pizza & Brewing Company w/ 2 others) http://t.
3.6060469E+017	2013-07-26 05:37:12	-116.979	32.6576	206505551	thadryan	Drinking a BPA (Belgian-Style Pale Ale) by @BreweryOmmegang @ Ryan's
3.6060321E+017	2013-07-26 05:31:19	-117.12033033	32.76351375	18792014	Earl100	I'm about this life homie they play this at the bar on #vhs ha ha ha @ Blind La
3.6059831E+017	2013-07-26 05:11:50	-117.203	32.8773	51834968	GottaTickemAll	Drinking a Stone Farking Wheaton w00stout by @StoneBrewingCo @ Beeral
3.6059609E+017	2013-07-26 05:03:01	-117.00593948	32.80650222	17745551	dmoralesf	Why does a #beer sound so good right now? #craftbeer #cycling @ 125 Bike
3.6059516E+017	2013-07-26 04:59:19	-117.34772801	33.15989132	27753347	Slamica	@J_bogert wants a pie (at @PizzaPortBeer) http://t.co/Ti8HSeZvul
3.6059499E+017	2013-07-26 04:58:39	-117.27004051	32.99278223	22410004	willkimbley	#CUErockstar Solana Beach #beerCUE comes with a firepit @ Chief's Burge
3.6059321E+017	2013-07-26 04:51:35	-117.12033033	32.76351375	18792014	Earl100	#vegan #pizza I'm about that life right now @ Blind Lady Ale House http://t.co
3.6059149E+017	2013-07-26 04:44:44	-117.252	32.7475	28627987	SeaJeb	Drinking an IPA by New English Brewing Co. at @newportpizza — http://t.co
3.6058680E+017	2013-07-26 04:26:07	-117.1777	32.90734	23131080	jessicatai	I'm at Green Flash Brewing Company - @greenflashbeer (San Diego
3.6058522E+017	2013-07-26 04:19:51	-117.21168508	32.74043584	22175554	CourtCFarrell	I'm at Stone Brewing World Bistro & Gardens Liberty Station - @stonebistro
3.6058053E+017	2013-07-26 04:01:11	-117.203	32.8773	51834968	GottaTickemAll	Drinking a Sculpin IPA by @bpbrewing @ Beeralot — http://t.co/kNMdH4qvz
3.6057821E+017	2013-07-26 03:51:59	-117.11760794	32.76348542	17632649	A7D	What #SummerNights are made for #livemusic #greenflashbrew #goodeats @
3.6057388E+017	2013-07-26 03:34:47	-117.23116	33.15229	1053472832	Michaelezz	Indian Joe Beer http://t.co/WIZ6nsnqdP

Figure 24: A sample extract of the dataset available for this project.

A total number of 26792 general beer- and craft beer related tweets were collected for this study in San Diego County during the 10 month time period. Out of the total number of tweets there are 7123 distinct Twitter users.

3.3 Qualitative Analysis

3.3.1 Goals and Purpose

The focus of this analysis is to analyze the textual content of the tweets. A major objective is to build an ontology which constitutes what is defined in this research as the ***microbrew theme***, consisting of a set of keywords which relate to craft beer. This means that only those tweets will be taken into account which have a relationship to craft beer and hence can be assigned to the microbrew theme.

Samples of tweets that will be filtered out:

" I need a beer"
"Beer, pizza and football #mykindofsaturday"
"free glass with beer, nice (@ The Duck Dive) [pic]: http://t.co/MH2wub2J"

For this research, only those tweets are considered as relevant which textual content closely relates to the microbrew theme. Samples include:

"Hanging at Helms Brewing. @HelmsBrewingCo. #craftbeer #chillin"
"Another micro brewery and craft beer bar. Love it. Porter beer is great! @ Pacific Beach Ale House"
"Handcrafted #beer degustation with dinner @ Karl Strauss Brewery & Restaurant"

The microbrew theme is made up of categories and subcategories such as ***Brewing Company Name***, ***Brand of Microbrew Beer***, ***Type of Microbrew Beer***, etc.

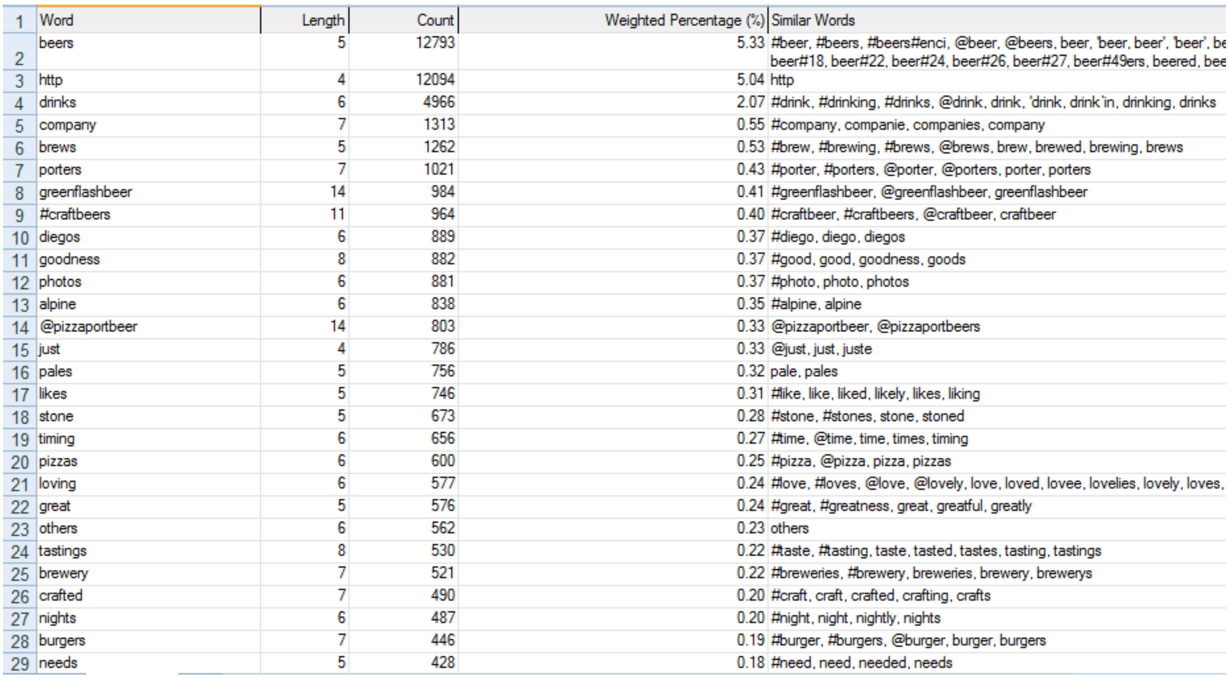
Each craft beer related tweet will be allocated to one or more categories.

Purposes of this qualitative analysis include:

1. Filter out the tweets that do not have an obvious relationship to the microbrew theme
2. Recognize named entities within the textual content of tweets
3. Categorize identified named entities within the textual content guided by an ontology-driven approach
4. Future usage: The ontology can be used as a fundament to analyze newly retrieved datasets fast and effectively

3.3.2 Word Extraction and Word Frequency

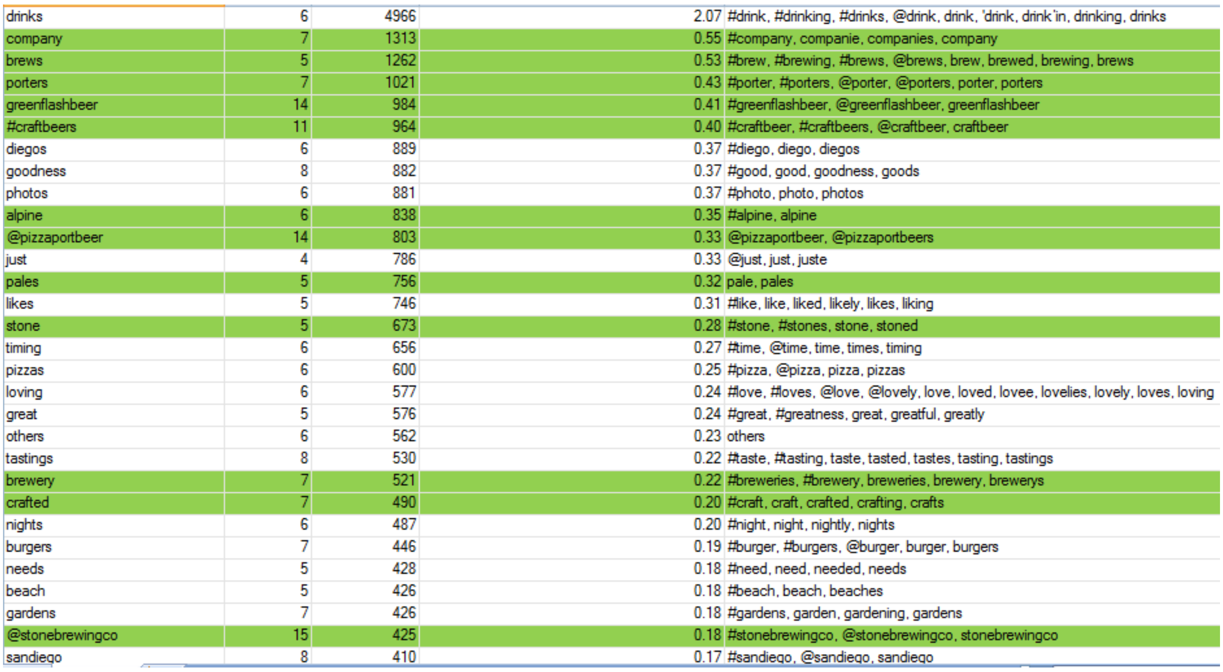
One of the very first steps of this qualitative analysis is to extract single words from the text of tweets and sort them by the number of occurrences in the entire dataset. For the Twitter dataset which is based on general beer related tweets, the query is run searching for the top 1000 words that have a minimum size of 4 characters. Figure 25 shows a list of the top 30 occurrences.



1	Word	Length	Count	Weighted Percentage (%)	Similar Words
2	beers	5	12793	5.33	#beer, #beers, #beers#enci, @beer, @beers, beer, 'beer, beer', 'beer', be beer#18, beer#22, beer#24, beer#26, beer#27, beer#49ers, beered, bee
3	http	4	12094	5.04	http
4	drinks	6	4966	2.07	#drink, #drinking, #drinks, @drink, drink, 'drink, drink'in, drinking, drinks
5	company	7	1313	0.55	#company, companie, companies, company
6	brews	5	1262	0.53	#brew, #brewing, #brews, @brews, brew, brewed, brewing, brews
7	porters	7	1021	0.43	#porter, #porters, @porter, @porters, porter, porters
8	greenflashbeer	14	984	0.41	#greenflashbeer, @greenflashbeer, greenflashbeer
9	#craftbeers	11	964	0.40	#craftbeer, #craftbeers, @craftbeer, craftbeer
10	diegos	6	889	0.37	#diego, diego, diegos
11	goodness	8	882	0.37	#good, good, goodness, goods
12	photos	6	881	0.37	#photo, photo, photos
13	alpine	6	838	0.35	#alpine, alpine
14	@pizzaportbeer	14	803	0.33	@pizzaportbeer, @pizzaportbeers
15	just	4	786	0.33	@just, just, juste
16	pales	5	756	0.32	pale, pales
17	likes	5	746	0.31	#like, like, liked, likely, likes, liking
18	stone	5	673	0.28	#stone, #stones, stone, stoned
19	timing	6	656	0.27	#time, @time, time, times, timing
20	pizzas	6	600	0.25	#pizza, @pizza, pizza, pizzas
21	loving	6	577	0.24	#love, #loves, @love, @lovely, love, loved, lovee, lovelies, lovely, loves,
22	great	5	576	0.24	#great, #greatness, great, greatful, greatly
23	others	6	562	0.23	others
24	tastings	8	530	0.22	#taste, #tasting, taste, tasted, tastes, tasting, tastings
25	brewery	7	521	0.22	#breweries, #brewery, breweries, brewery, brewerys
26	crafted	7	490	0.20	#craft, craft, crafted, crafting, crafts
27	nights	6	487	0.20	#night, night, nightly, nights
28	burgers	7	446	0.19	#burger, #burgers, @burger, burger, burgers
29	needs	5	428	0.18	#need, need, needed, needs

***Figure 25:** A list of the top 30 occurrences of keywords within the textual content of tweets. The "Count" column highlights the number of occurrences.*

Some words in the list above have or might have a strong craft beer relationship. Figure 26 marks some of the top words which are or might be closely related to the microbrew theme:



drinks	6	4966	2.07	#drink, #drinking, #drinks, @drink, drink, 'drink, drink'in, drinking, drinks
company	7	1313	0.55	#company, companie, companies, company
brews	5	1262	0.53	#brew, #brewing, #brews, @brews, brew, brewed, brewing, brews
porters	7	1021	0.43	#porter, #porters, @porter, @porters, porter, porters
greenflashbeer	14	984	0.41	#greenflashbeer, @greenflashbeer, greenflashbeer
#craftbeers	11	964	0.40	#craftbeer, #craftbeers, @craftbeer, craftbeer
diegos	6	889	0.37	#diego, diego, diegos
goodness	8	882	0.37	#good, good, goodness, goods
photos	6	881	0.37	#photo, photo, photos
alpine	6	838	0.35	#alpine, alpine
@pizzaportbeer	14	803	0.33	@pizzaportbeer, @pizzaportbeers
just	4	786	0.33	@just, just, juste
pales	5	756	0.32	pale, pales
likes	5	746	0.31	#like, like, liked, likely, likes, liking
stone	5	673	0.28	#stone, #stones, stone, stoned
timing	6	656	0.27	#time, @time, time, times, timing
pizzas	6	600	0.25	#pizza, @pizza, pizza, pizzas
loving	6	577	0.24	#love, #loves, @love, @lovely, love, loved, lovee, lovelies, lovely, loves, loving
great	5	576	0.24	#great, #greatness, great, greatful, greatly
others	6	562	0.23	others
tastings	8	530	0.22	#taste, #tasting, taste, tasted, tastes, tasting, tastings
brewery	7	521	0.22	#breweries, #brewery, breweries, brewery, brewerys
crafted	7	490	0.20	#craft, craft, crafted, crafting, crafts
nights	6	487	0.20	#night, night, nightly, nights
burgers	7	446	0.19	#burger, #burgers, @burger, burger, burgers
needs	5	428	0.18	#need, need, needed, needs
beach	5	426	0.18	#beach, beach, beaches
gardens	7	426	0.18	#gardens, garden, gardening, gardens
@stonebrewingco	15	425	0.18	#stonebrewingco, @stonebrewingco, stonebrewingco
sandiego	8	410	0.17	#sandiego, @sandiego, sandiego

Figure 26: A list of the top keywords which relate or might relate to craft beer (green rows).

Whether a word is related to the microbrew theme or not is determined ad-hoc by the interpretation of the word. Examples of such words include:

- craftbeer → A beer that is not brewed by mass-production beer companies, but rather by local breweries (microbreweries)
- societe → A micro brewing company in San Diego
- brewpub → General term for a pub selling microbrewery products

3.3.3 Named Entity Recognition

After extracting words which are highly related to craft beer, the next step is to identify named entities in preparation for building the ontology.

This process consists of assigning the previously extracted words to the respective category.

Some words need to concatenate one or more words to have a specific meaning. For example the word **Karl** is a male first name and doesn't have a specific craft beer related meaning by nature. When concatenated with the word **Strauss** it becomes → **Karl Strauss**, which is one of the largest and most famous micro brewing companies in San Diego County.

Named entities identified out of the Twitter dataset are referred to as ***Twitter Named Entities (TNE)***. They correspond to atomic elements within the textual content of tweets constituted by the user following a general linguistic usage.

Here's a list of some of the identified TNEs which constitute the micro brew theme:

Twitter Name Entities (Examples)
#craftbeer
#craftbeerbliss
#PizzaPort
@callahanspub
@greenflashbeer
@pizzaportbeer
@sdbeerco
@StoneBrewingCo
@taproom
Alesmith Brewing Company
Alpine Beer Company
Beech Street IPA
Beer tasting
Black Orchid Stout
Bottlecraft (Beer Shop and Tasting Room)
Coronado Brewing Company @toronadosd
East Village Pilsner
Fezziwig's Brewing Company
Green Flash Brewing Company
Hillcrest Brewing Company
IPA
Jasper Extra Pale Ale
Junk In Da Trunkel
Karl Strauss Brewery & Restaurant
Lost Abbey's
Mad River Brewing Company
Manzanita Brewing Company
Monkey Paw Brewing Company
Navel College Old Porter
Offbeat Brewing Company
Oggi's Pizza & Brewing Company

3.3.4 Micro Brew Ontology

Based on the previously defined TNE, an ontology is developed which assigns each entity to a certain category. The ontology consists of three major categories:

1. **Breweries** – Consists of subcategories representing each brewery
2. **Brewery Amenities/Services** – Facilities which breweries offer such as tasting room or beer shop
3. **Beer Products** – The beer products which breweries offer categorized in types, styles and brands of beer

Figure 27 shows an extract of the visualization of the categories of the overall ontology.

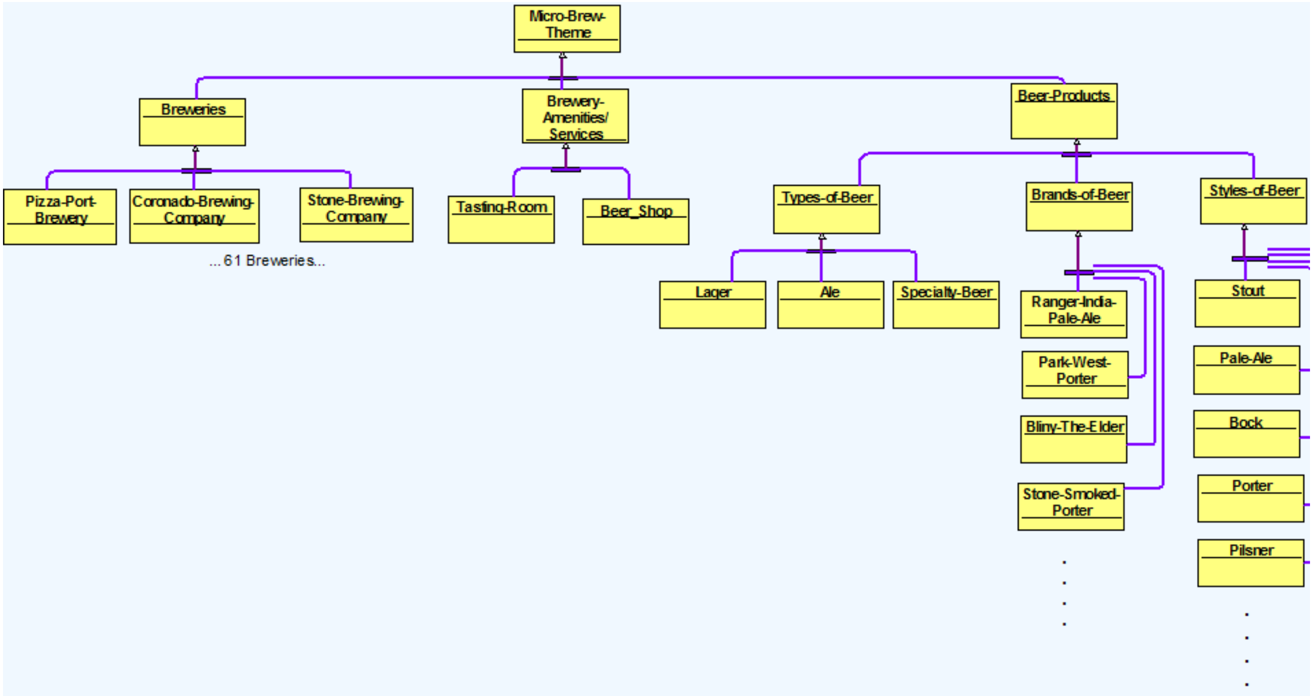


Figure 27: The overall craft beer ontology.

The named entities identified in the textual content of the tweets correspond to the instances (individuals) of the ontology. Each brewery within the Breweries category has different occurrences of named entities within the textual content of tweets. As an example, people mention ***Stone Brewing Company*** in tweets as follows:

- @StoneBrewingCo
- #StoneBrewing
- Stone Brewing Company
- Stone Brewery

Figure 28 shows an extract of the visualization of the Breweries category with their respective TNE.

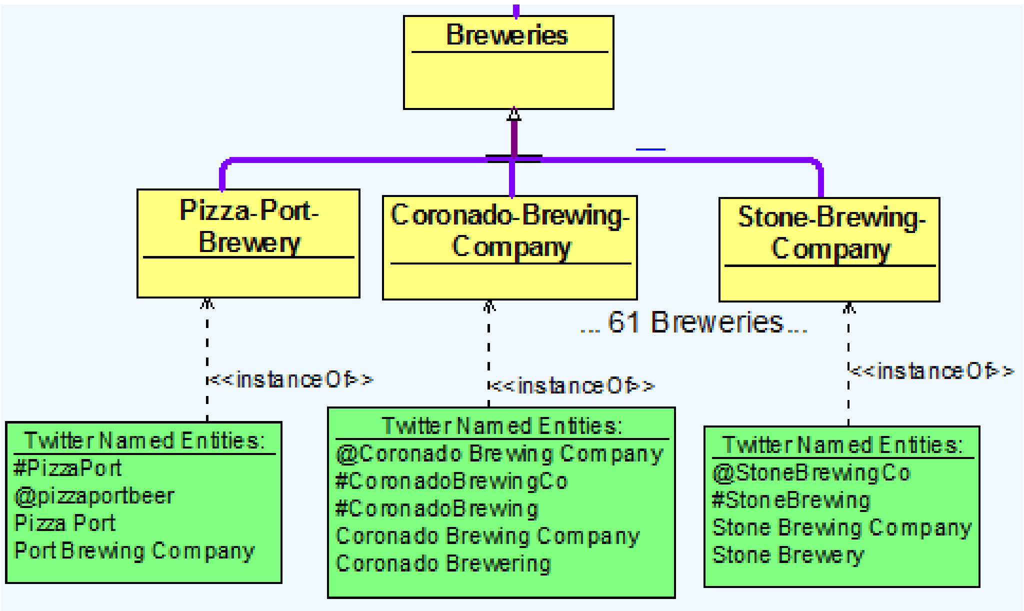


Figure 28: *A closer look at the Breweries category showing all the respective TNEs (green).*

To complete the ontology, relationships between the categories can be defined.

As an example, in the tweet

"Stone brewery serves THE best beers! I love Stone Smoked Porter."

two categories can be identified:

1. The TNE: 'Stone Brewery' → belonging to the ***Stone-Brewery-Company*** category
2. The TNE: 'Stone Smoked Porter' → belonging to the category ***Brands-of-Beer*** category

The definition of the general relationship would be defined as:

Brewery **serves** (1...n) Brands-of-Beers

For this study, modeling of relationships is not taken into account. This ontology should only serve as groundwork for further research!

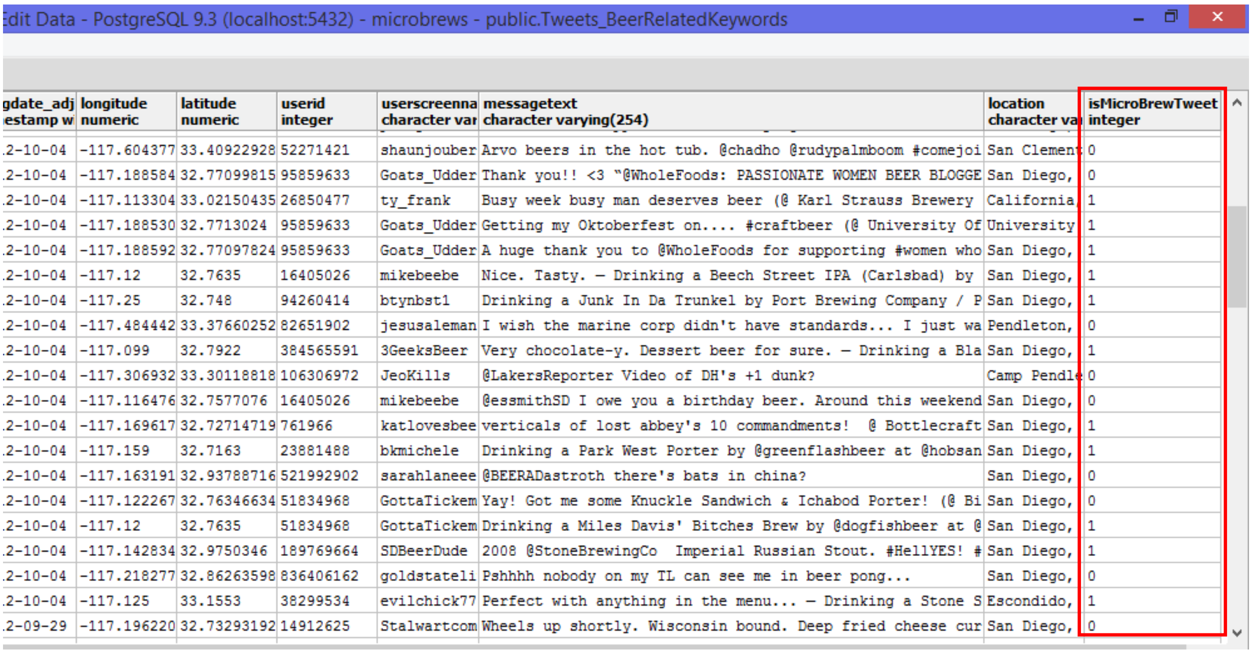
This minimum ontology serves the following purposes of this study:

1. filter out tweets that cannot be associated with the microbrew theme,
2. identify named entities within the textual content of the tweets and
3. identify tweets that obviously relate to the microbrew theme for further analysis
4. allocate tweets to the respective breweries if such information is available within the textual content of the tweets.

3.3.5 Identify craft beer related tweets

Based on the previously developed ontology each tweet is either assigned to the micro brew theme or not (when no micro brew related named entity is found within the tweet text). Only the tweets relating to craft beer are taken into account for further analysis. This is accomplished by adding an additional database field to the Twitter dataset (*isMicroBrewTweets*). A script assigns each tweet either the value 1 (relation to craft beer) or the value 0 (no relation to craft beer).

An extract of the respective data table is shown in Figure 29.



:dit Data - PostgreSQL 9.3 (localhost:5432) - microbrews - public.Tweets_BeerRelatedKeywords

gdate_adj estamp w	longitude numeric	latitude numeric	userid integer	userscreenna character var	messagetext character varying(254)	location character va	isMicroBrewTweet integer
.2-10-04	-117.604377	33.40922928	52271421	shaunjouber	Arvo beers in the hot tub. @chadho @rudypalmboom #comejoi	San Clement	0
.2-10-04	-117.188584	32.77099815	95859633	Goats_Udder	Thank you!! <3 "@WholeFoods: PASSIONATE WOMEN BEER BLOGGE	San Diego,	0
.2-10-04	-117.113304	33.02150435	26850477	ty_frank	Busy week busy man deserves beer (@ Karl Strauss Brewery	California,	1
.2-10-04	-117.188530	32.7713024	95859633	Goats_Udder	Getting my Oktoberfest on.... #craftbeer (@ University Of	University	1
.2-10-04	-117.188592	32.77097824	95859633	Goats_Udder	A huge thank you to @WholeFoods for supporting #women who	San Diego,	1
.2-10-04	-117.12	32.7635	16405026	mikebeebe	Nice. Tasty. – Drinking a Beech Street IPA (Carlsbad) by	San Diego,	1
.2-10-04	-117.25	32.748	94260414	btynbst1	Drinking a Junk In Da Trunkel by Port Brewing Company / P	San Diego,	1
.2-10-04	-117.484442	33.37660252	82651902	jesusaleman	I wish the marine corp didn't have standards... I just wa	Pendleton,	0
.2-10-04	-117.099	32.7922	384565591	3GeeksBeer	Very chocolate-y. Dessert beer for sure. – Drinking a Bla	San Diego,	1
.2-10-04	-117.306932	33.30118818	106306972	JeoKills	@LakersReporter Video of DH's +1 dunk?	Camp Pendle	0
.2-10-04	-117.116476	32.7577076	16405026	mikebeebe	@essmithSD I owe you a birthday beer. Around this weekend	San Diego,	0
.2-10-04	-117.169617	32.72714719	761966	katlovesbee	verticals of lost abbey's 10 commandments! @ Bottlecraft	San Diego,	1
.2-10-04	-117.159	32.7163	23881488	bkmichele	Drinking a Park West Porter by @greenflashbeer at @hobsan	San Diego,	1
.2-10-04	-117.163191	32.93788716	521992902	sarahlaneee	@BEERADastroth there's bats in china?	San Diego,	0
.2-10-04	-117.122267	32.76346634	51834968	GottaTickem	Yay! Got me some Knuckle Sandwich & Ichabod Porter! (@ Bi	San Diego,	0
.2-10-04	-117.12	32.7635	51834968	GottaTickem	Drinking a Miles Davis' Bitches Brew by @dogfishbeer at @	San Diego,	1
.2-10-04	-117.142834	32.9750346	189769664	SDBeerDude	2008 @StoneBrewingCo Imperial Russian Stout. #HellYES! #	San Diego,	1
.2-10-04	-117.218277	32.86263598	836406162	goldstateli	Pshhhh nobody on my TL can see me in beer pong...	San Diego,	0
.2-10-04	-117.125	33.1553	38299534	evilchick77	Perfect with anything in the menu... – Drinking a Stone S	Escondido,	1
.2-09-29	-117.196220	32.73293192	14912625	Stalwartcom	Wheels up shortly. Wisconsin bound. Deep fried cheese cur	San Diego,	0

Figure 29: *An additional field was added to the data model in order to differentiate craft beer related tweet from non- craft beer related ones.*

3.4 Spatial Analysis

In this section a spatial analysis is conducted on those tweets which were previously classified as craft beer related tweets. This includes the creation of a heat map showing the overall distribution of craft beer related tweets; analyze tweet densities on the neighborhood and ZIP code level, and the analysis of tweet quantities for existing breweries.

3.4.1 Hotspot Map (Kernel Density Function)

A raster-based hotspot map based on the kernel density function is generated to obtain an overall picture of the distribution of all microbrew related tweets (see Figure 9). This map might be an indicator which areas hold a high popularity towards craft beer.

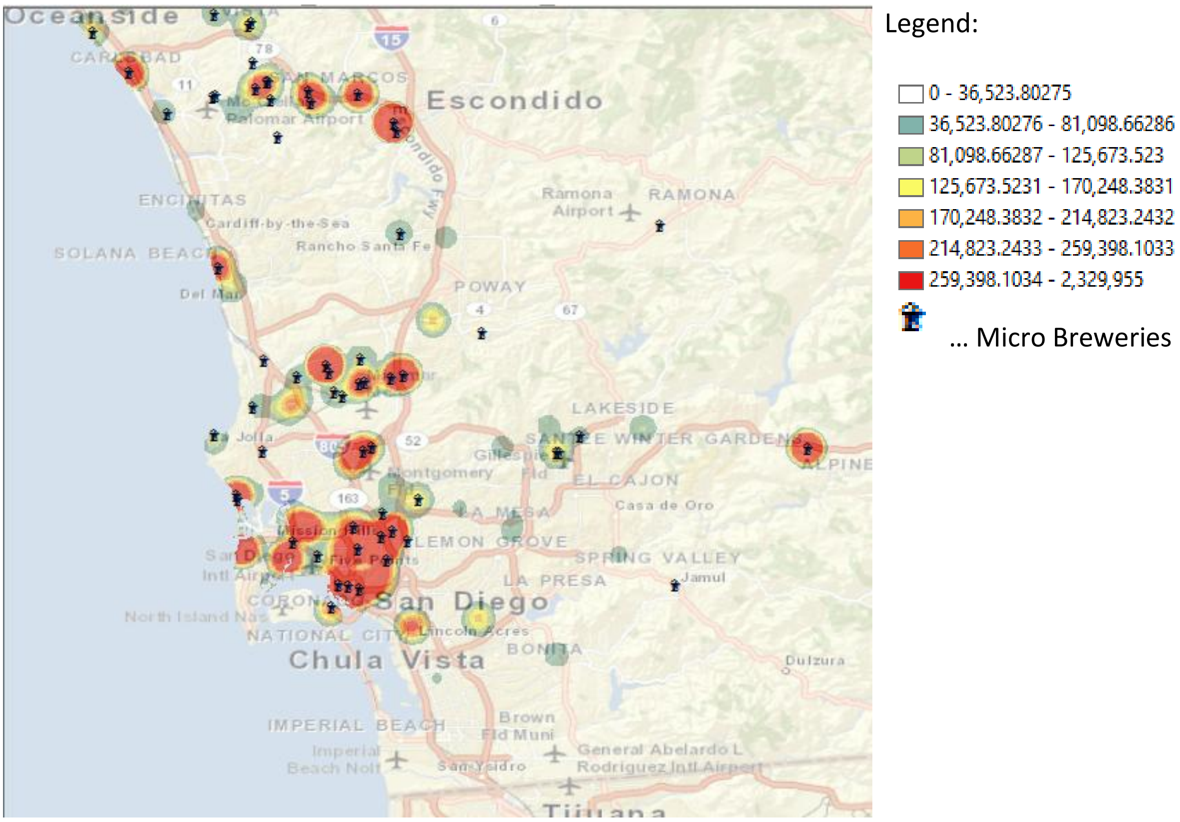
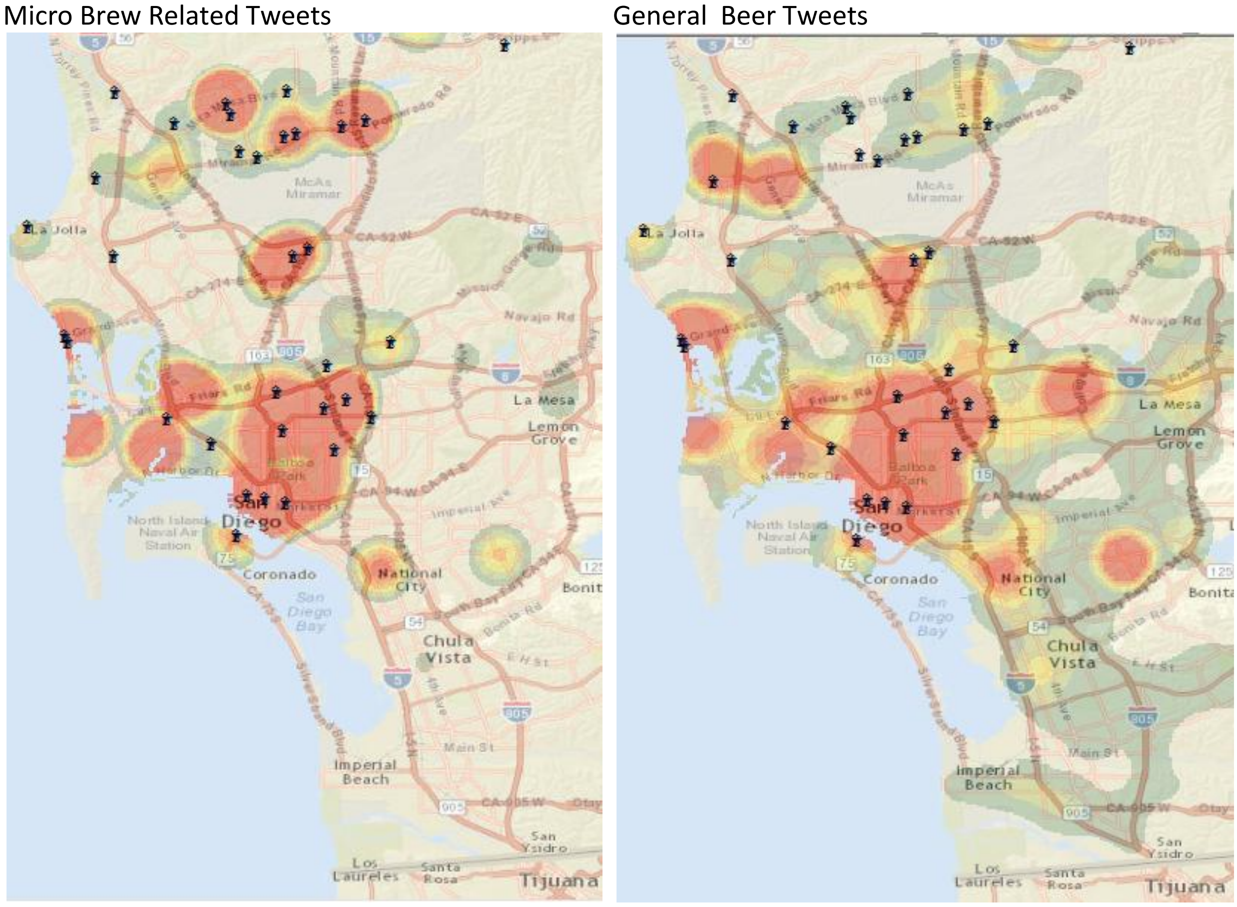


Figure 30: *A heat map displaying the distribution of craft beer related tweets.The red areas indicate a high tweet density. ½ Standard Deviation is chosen as the classifcation method for this map.*

Furthermore the densities of tweets with craft beer keywords and general beer related keywords were compared in Figure 31. The results show that the tweets containing general beer related keywords are widely spread out compared to tweets with craft beer related keywords.



Micro Brew Related Tweets

General Beer Tweets

Figure 31: *Comparison of the spatial distribution of craft beer related and general beer related tweets.*

3.4.2 Tweet Densities – Neighboorhoods City of San Diego

The densities of craft beer related tweets are calculated for each neighborhood of the city of San Diego and visualized on a choropleth map showing the number of tweets per neighborhood (see Figures 32-34). This was accomplished by applying the Spatial Join operation.

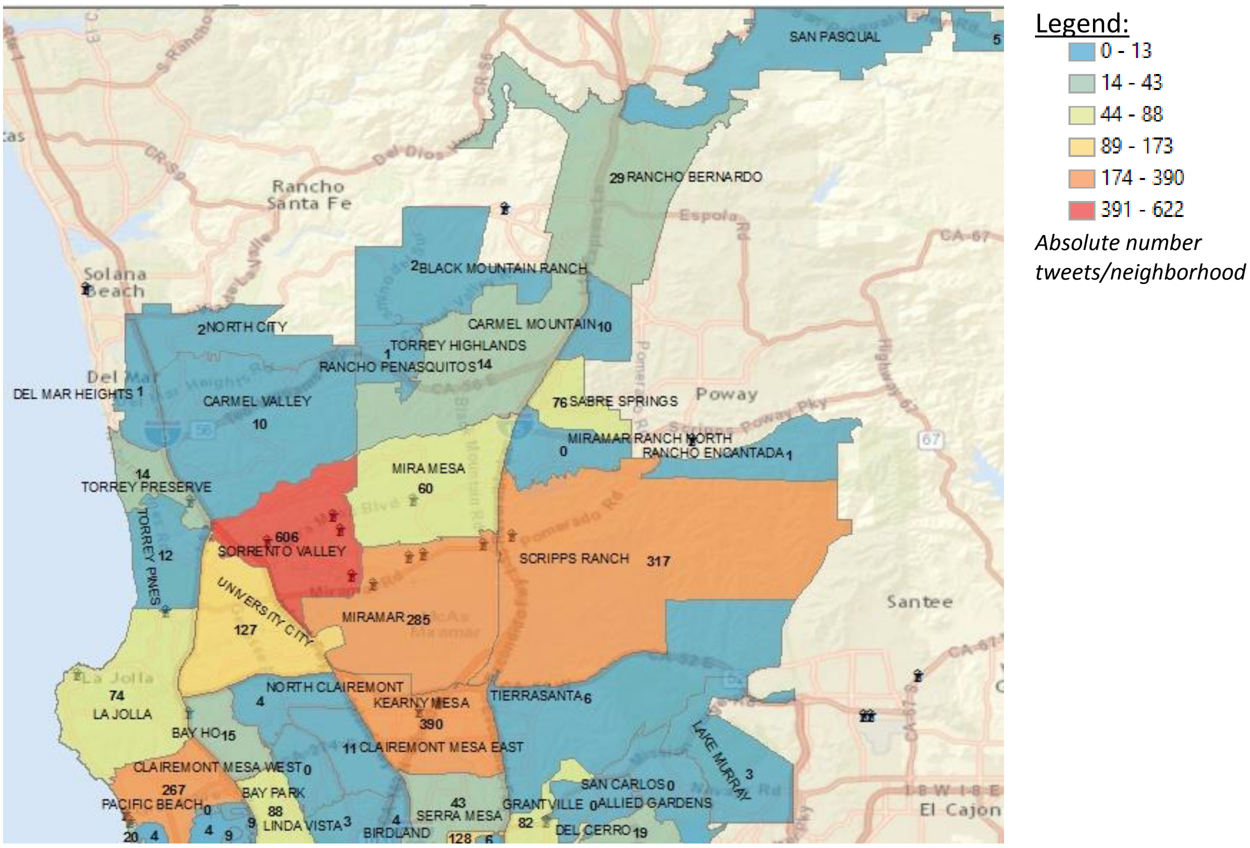


***Figure 32:** Absolute number of tweets per neighborhood – northern part of the city of San Diego.*

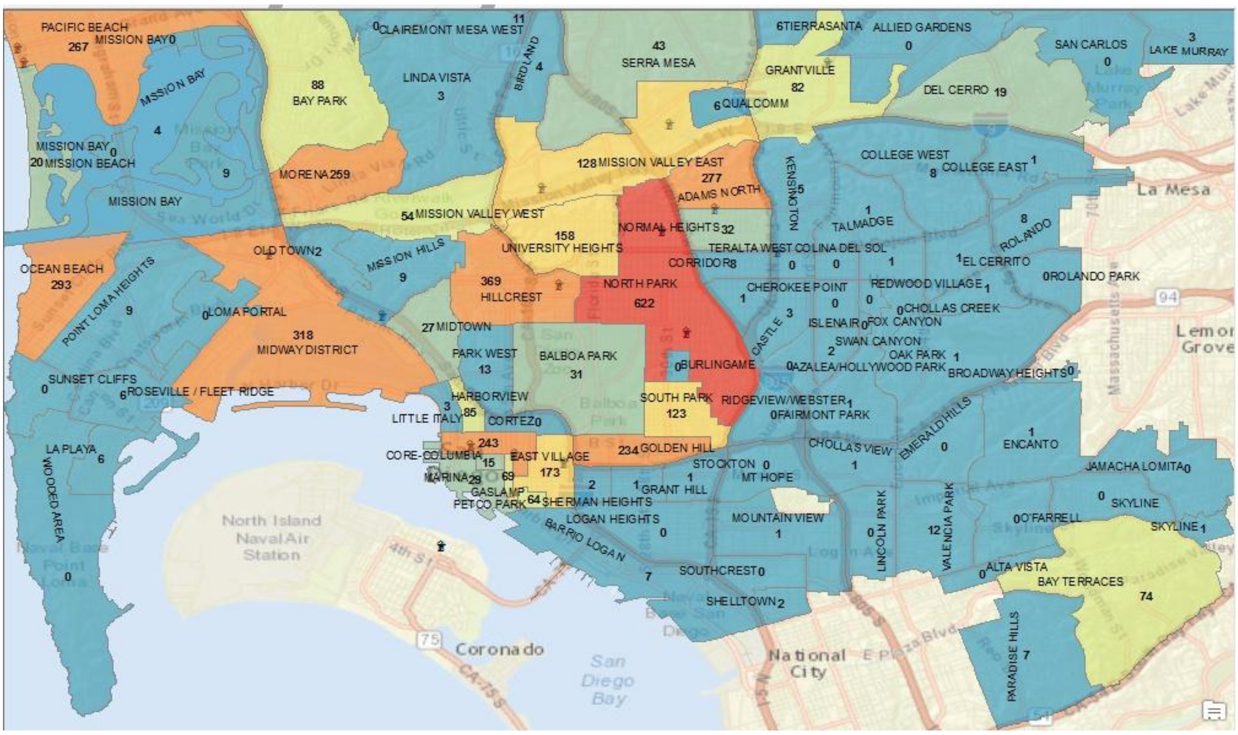


***Figure 33:** Absolute number of tweets per neighborhood – Downtown area and surroundings of San Diego.*

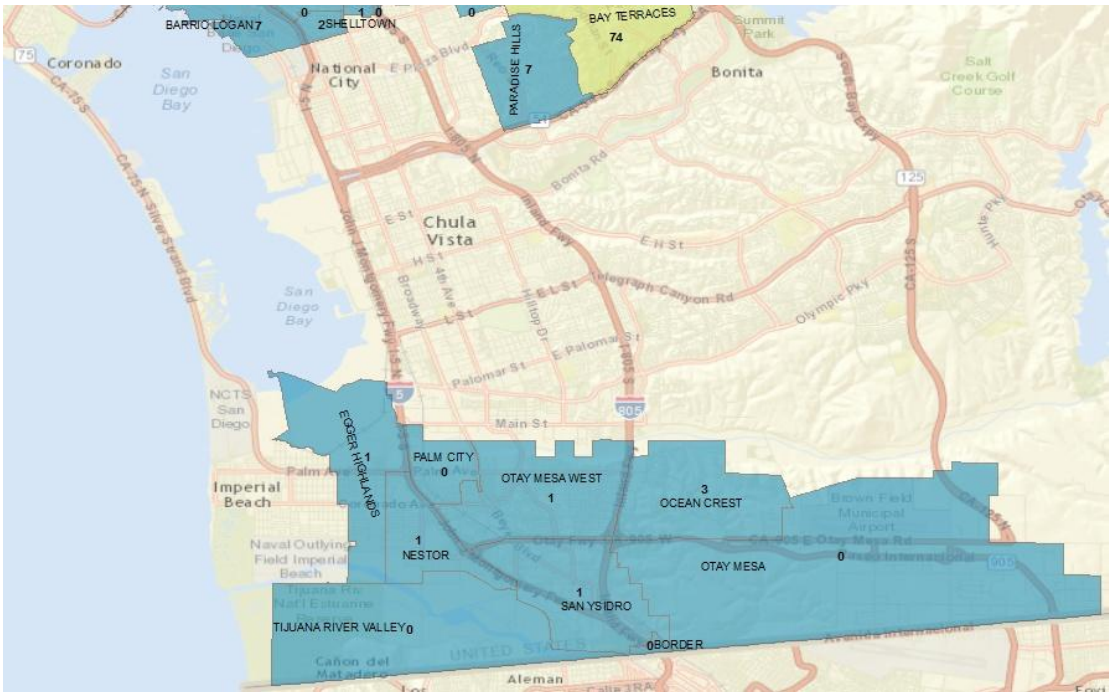


Figure 34: Absolute number of tweets per neighborhood - southern part of the city of San Diego.

The following table shows the top 40 neighborhoods of the city of San Diego regarding absolute number of tweets:

Rank	Neighborhood Name	Absolute number of tweets
1	NORTH PARK	622
2	SORRENTO VALLEY	606
3	KEARNY MESA	390
4	HILLCREST	369
5	MIDWAY DISTRICT	318
6	SCRIPPS RANCH	317
7	OCEAN BEACH	293
8	MIRAMAR	285
9	ADAMS NORTH	277
10	PACIFIC BEACH	267
11	MORENA	259
12	CORE-COLUMBIA	243
13	GOLDEN HILL	234
14	EAST VILLAGE	173
15	UNIVERSITY HEIGHTS	158
16	MISSION VALLEY EAST	128
17	UNIVERSITY CITY	127
18	SOUTH PARK	123
19	BAY PARK	88
20	LITTLE ITALY	85
21	GRANTVILLE	82
22	SABRE SPRINGS	76

23	BAY TERRACES	74
24	LA JOLLA	74
25	GASLAMP	69
26	PETCO PARK	64
27	MIRA MESA	60
28	MISSION VALLEY WEST	54
29	SERRA MESA	43
30	NORMAL HEIGHTS	32
31	BALBOA PARK	31
32	MARINA	29
33	RANCHO BERNARDO	29
34	MIDTOWN	27
35	MISSION BEACH	20
36	DEL CERRO	19
37	HORTON PLAZA	15
38	BAY HO	15
39	RANCHO PENASQUITOS	14
40	TORREY PRESERVE	14

Table 4: Number of craft beer related tweets per neighborhood.

3.4.3 Tweet Densities – ZIP Code scale San Diego County

In this subsection, the densities of tweets that relate to craft beer are calculated on the 5-digit ZIP code level and visualized on a choropleth map (see Figure 35 and Figure 36).

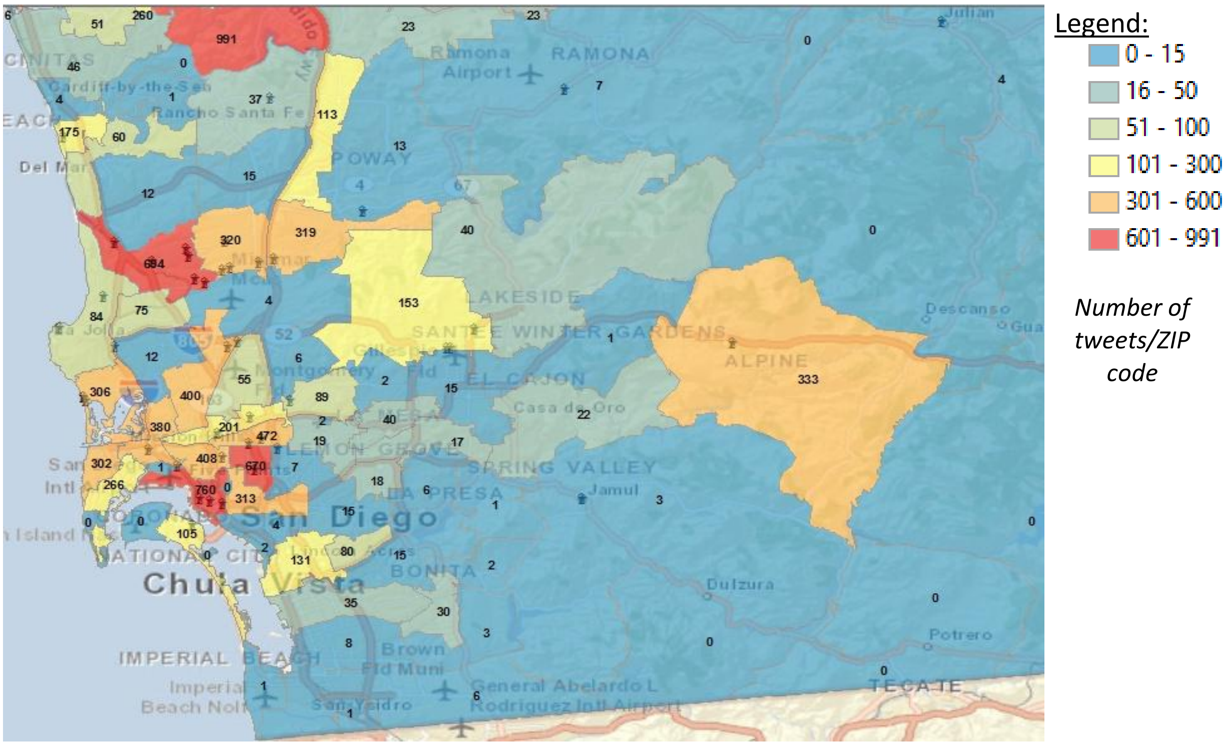


Figure 35: *Absolute number craft beer related tweets per ZIP Code – San Diego South County.*

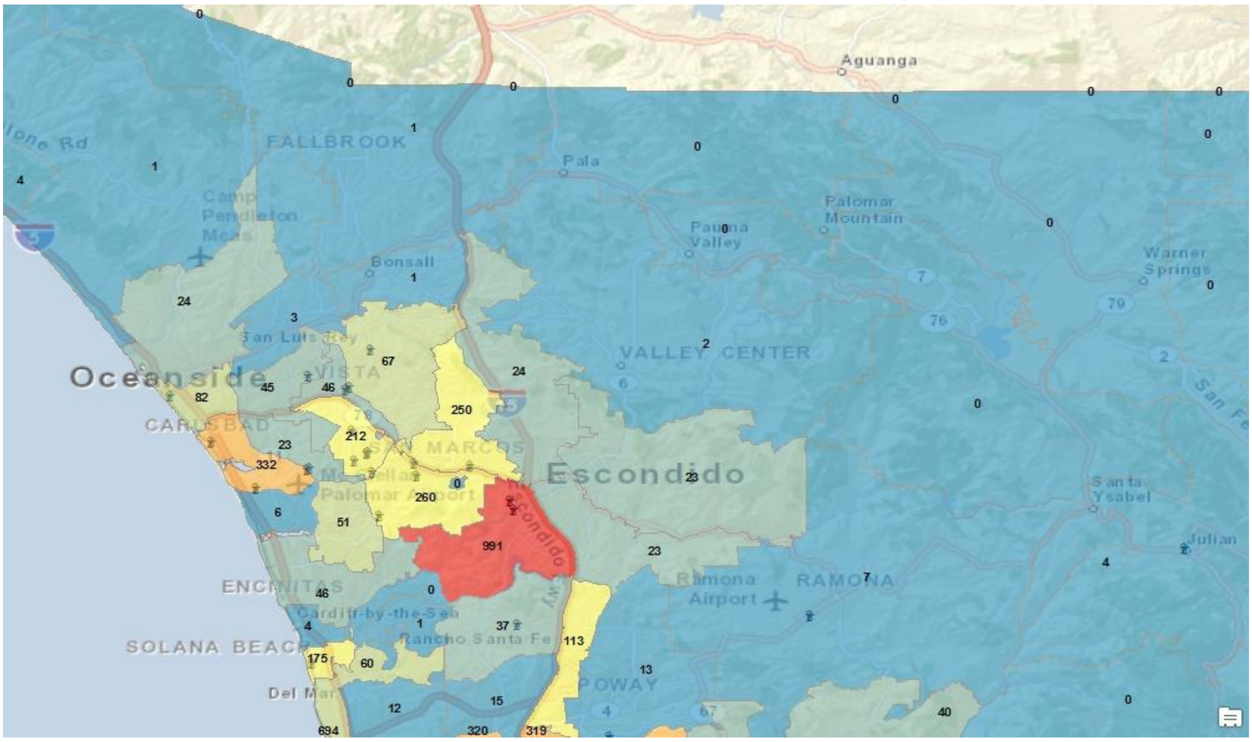


Figure 36: *Absolute number craft beer related tweets per ZIP Code – San Diego South County.*

The following list shows the top 40 ZIP code areas regarding the absolute number of tweets:

Rank	ZIP Code	City Name	Absolute number of tweets
1	92029	Escondido	991
2	92101	San Diego	760
3	92121	San Diego	694
4	92104	San Diego	670
5	92116	San Diego	472
6	92103	San Diego	408
7	92111	San Diego	400
8	92110	San Diego	380
9	91901	Alpine	333
10	92008	Carlsbad	332
11	92126	San Diego	320
12	92131	San Diego	319
13	92102	San Diego	313
14	92109	San Diego	306
15	92107	San Diego	302
16	92106	San Diego	266
17	92078	San Marcos	260
18	92069	San Marcos	250
19	92081	Vista	212
20	92108	San Diego	201
21	92075	Solana Beach	175
22	92071	Santee	153

23	91950	National City	131
24	92128	San Diego	113
25	92118	Coronado	105
26	92120	San Diego	89
27	92037	La Jolla	84
28	92054	Oceanside	82
29	92139	San Diego	80
30	92122	San Diego	75
31	92084	Vista	67
32	92014	Del Mar	60
33	92123	San Diego	55
34	92009	Carlsbad	51
35	92083	Vista	46
36	92024	Encinitas	46
37	92056	Oceanside	45
38	92040	Lakeside	40
39	91942	La Mesa	40
40	92127	San Diego	37

Figure 37: *Number of craft beer related tweets per ZIP code.*

3.4.4 Tweet – Brewery Allocation

All the craft beer related tweets which have a connection to a certain brewery were allocated to the respective breweries based on the previously developed ontology.

The data is visualized on a map as graduated symbols depicting the location of the breweries (see Figures 38-41). The symbol size of each brewery relates to the number of tweets per brewery. The quantity inside the symbol indicates the absolute number of tweets per brewery.

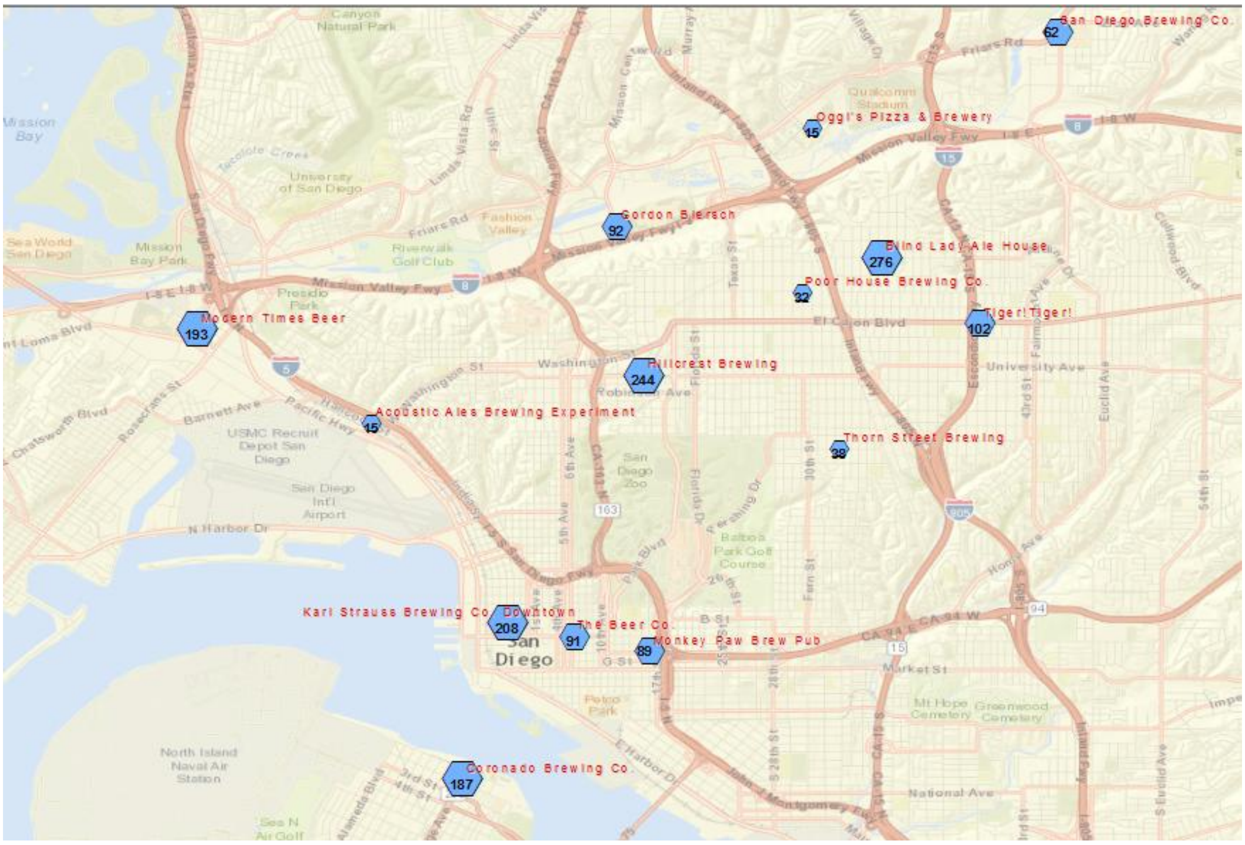


Figure 38: *Number of tweets per brewery displayed as graduated symbols – Downtown San Diego.*

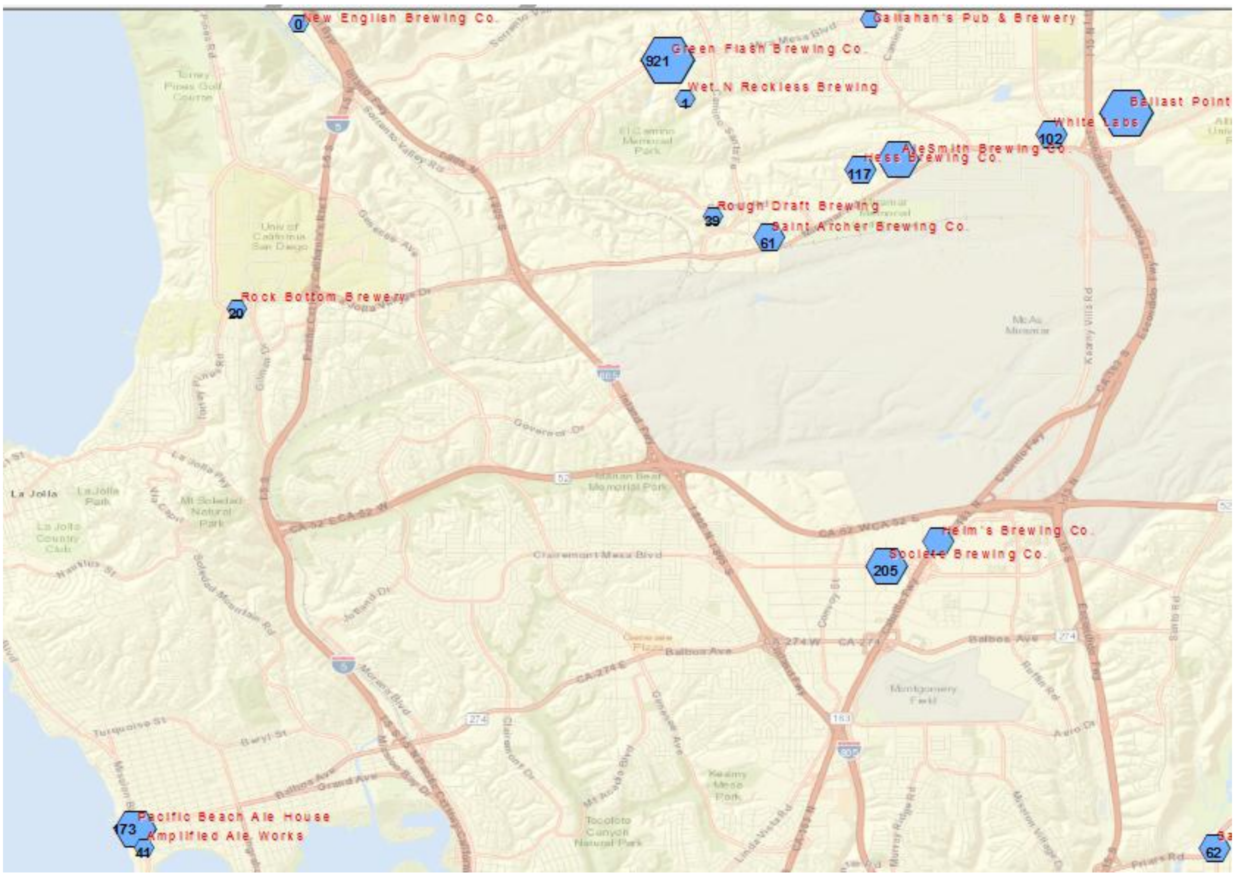


Figure 39: *Number of tweets per brewery displayed as graduated symbols – Mid-Western part of San Diego County.*

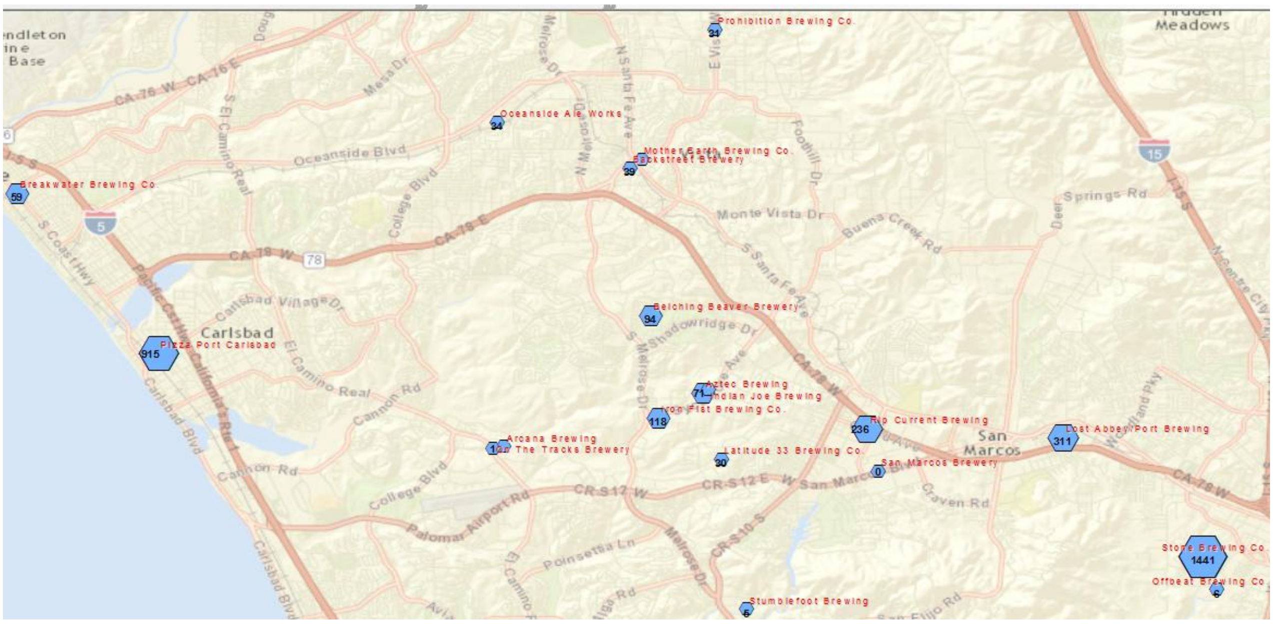


Figure 40: *Number of tweets per brewery displayed as graduated symbols – Northern part of San Diego County.*

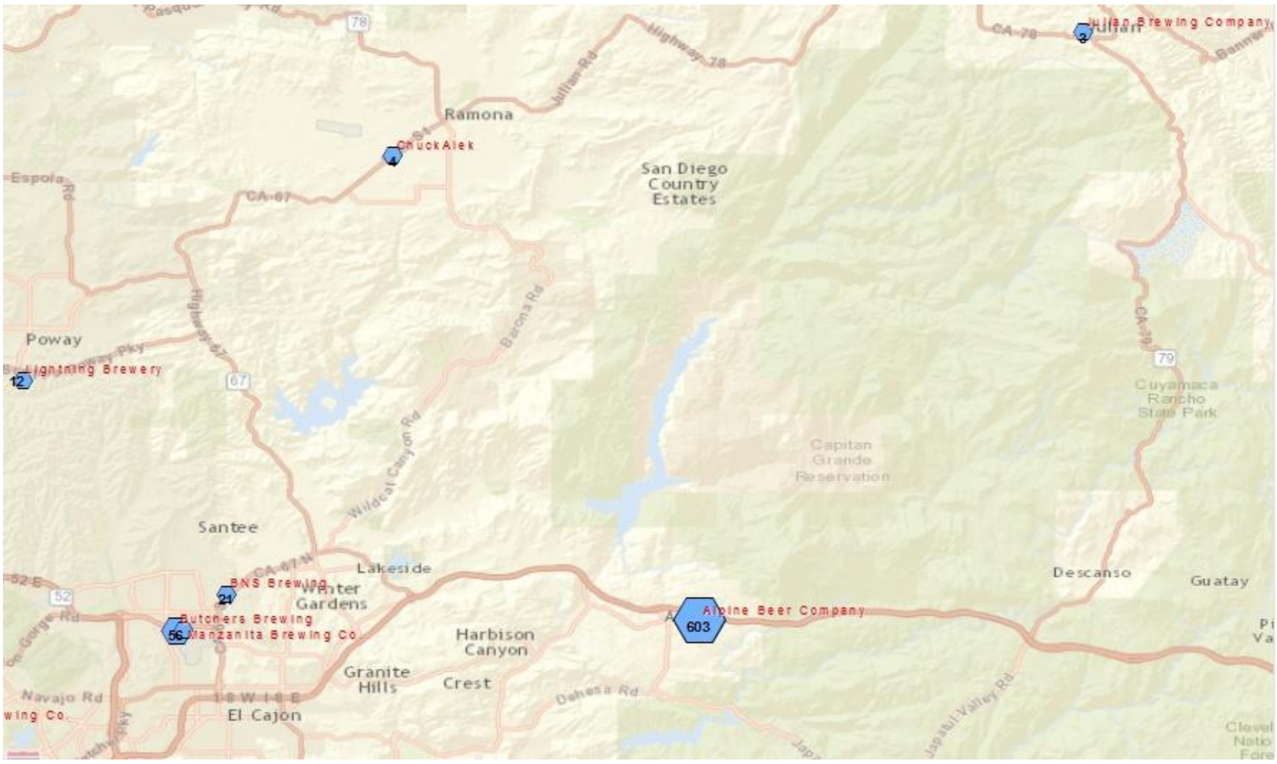
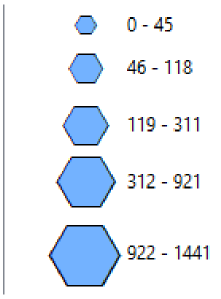


Figure 41: *Number of tweets per brewery displayed as graduated symbols – Eastern part of San Diego County.*



Number of tweets per brewery

3.5 Quantitative Analysis

This section focuses on the analysis of the demographic characteristics of craft beer consumers such as gender, income, age, etc.

3.5.1 Consumer Data

A sample of a 100 users frequently posting craft beer related tweets was randomly selected and their public Twitter profile linked to the public lifestyle database ReferenceUSA to retrieve consumer information regarding the following demographic variables:

- Gender
- Age
- Income
- Marital Status
- Religion
- Race

Privacy Statement: In order to protect the privacy of the users, personal information such as user id or username was removed from the sample.

3.5.2 Demographic Variables

This subsection delivers an exploratory analysis of each consumer variable.

3.5.2.1 Gender

Out of the 100 users selected out of the entire set of tweets, 30% were females and 70 % were males (see Figure 42).

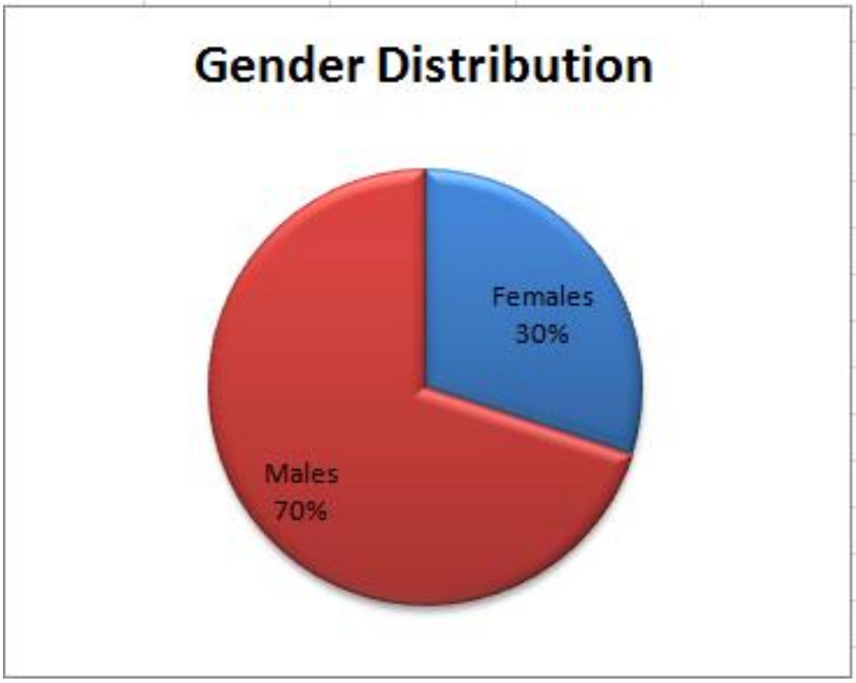


Figure 42*: According to the sample 70% of craft beer consumers are male whereas 30% are female.*

3.5.2.2 Age

The predominant craft beer consumer age group is 25 -34 years (about 37 % of the consumers). Figure 43 classifies the age of craft beer consumers in eight groups and the results are shown in a histogram.

Statistical consumer age values:
Mean Age: 35
Median Age: 35
Standard Deviation: +- 9.4

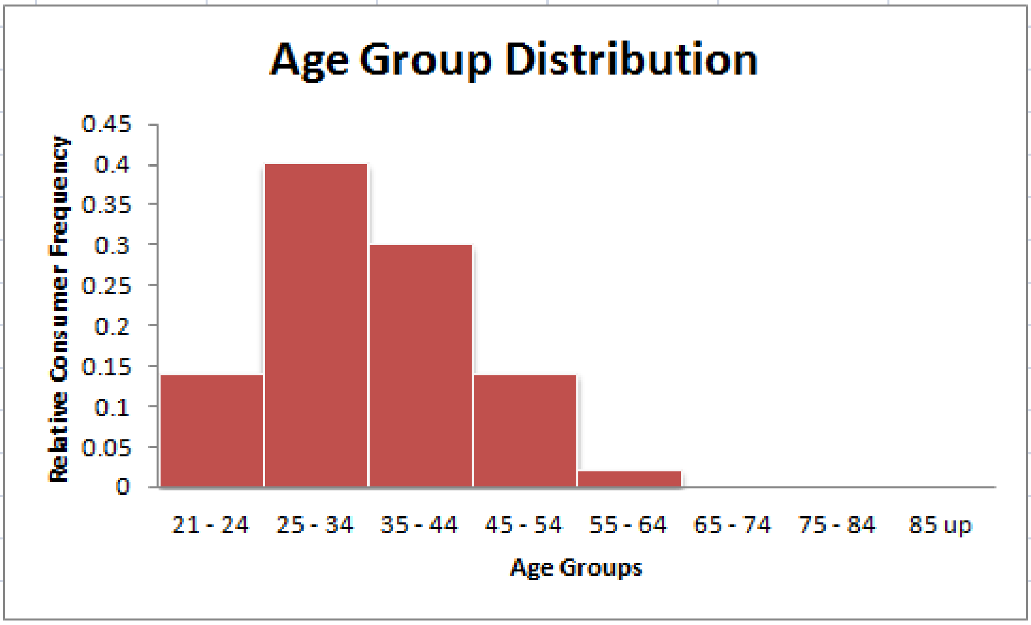


Figure 43: *A histogram showing the age distribution of craft beer consumers. The predominant age group of craft beer consumers is 25-34.*

The top five age groups are further summarized in a pie chart (see Figure 44).

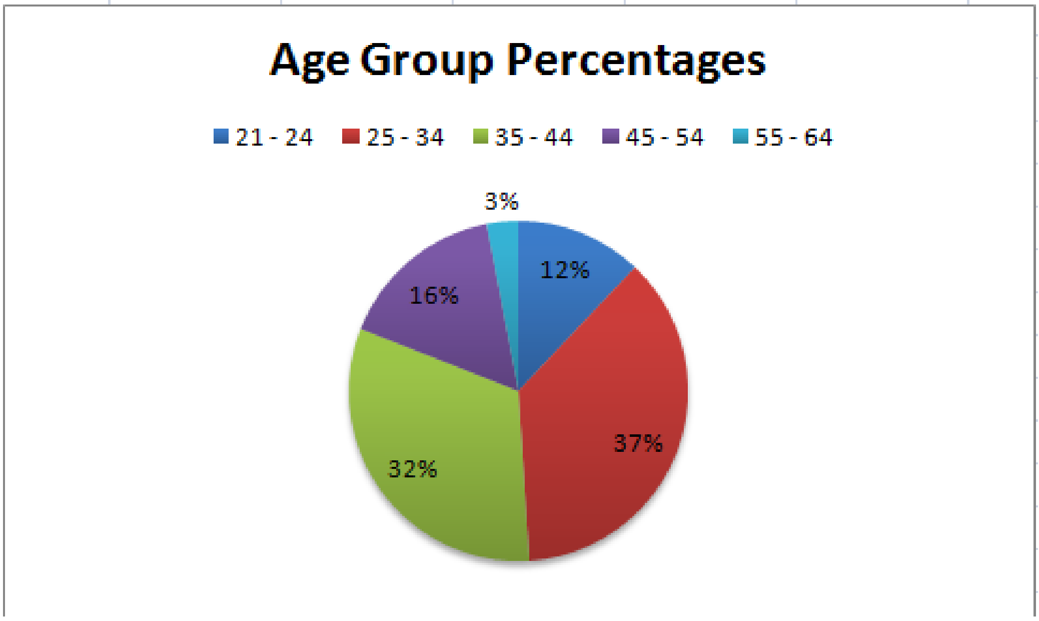


Figure 44: *Pie chart highlighting the distribution of the top five age groups of craft beer consumers.*

3.5.2.3 Income

The leading income group of the sample of craft beer consumers is 90.000$ - 150.000$ annually (about 25 % of the consumers). Figure 45 classifies the income of craft beer consumers in seven groups and the results are shown in a histogram.

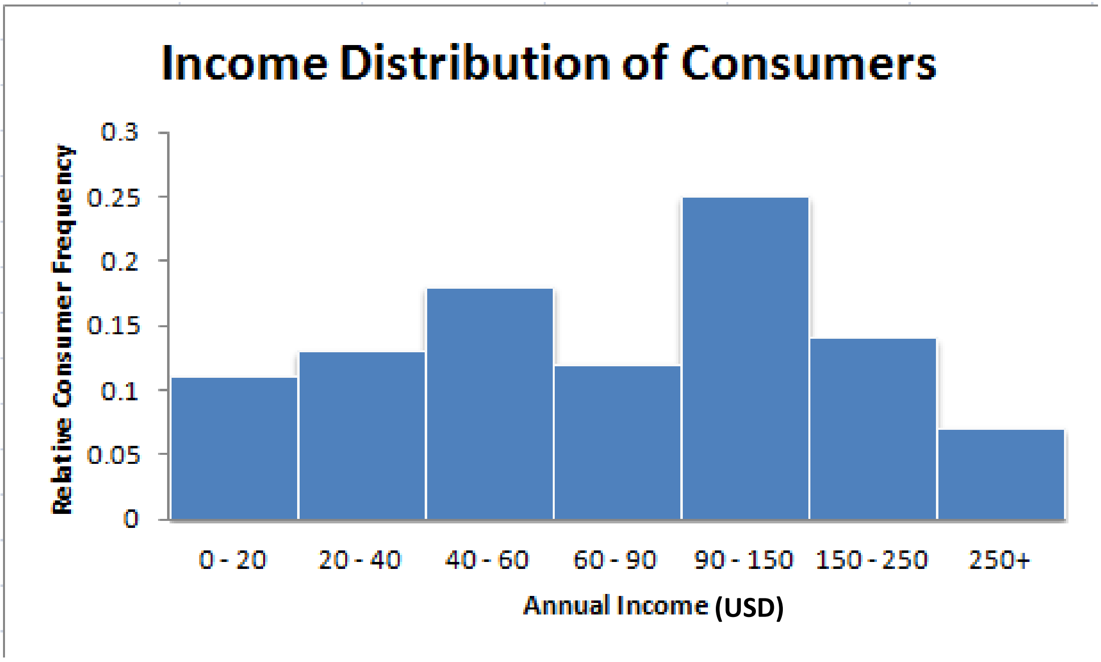


Figure 45: *A histogram displaying the craft beer consumer proportions for each income group (in US $). The predominating number of craft consumers (25%) has an income between $90k and $150k.*

3.5.2.4 Marital Status

The group of married consumers slightly prevails over the group of single consumers (see Figure 46).

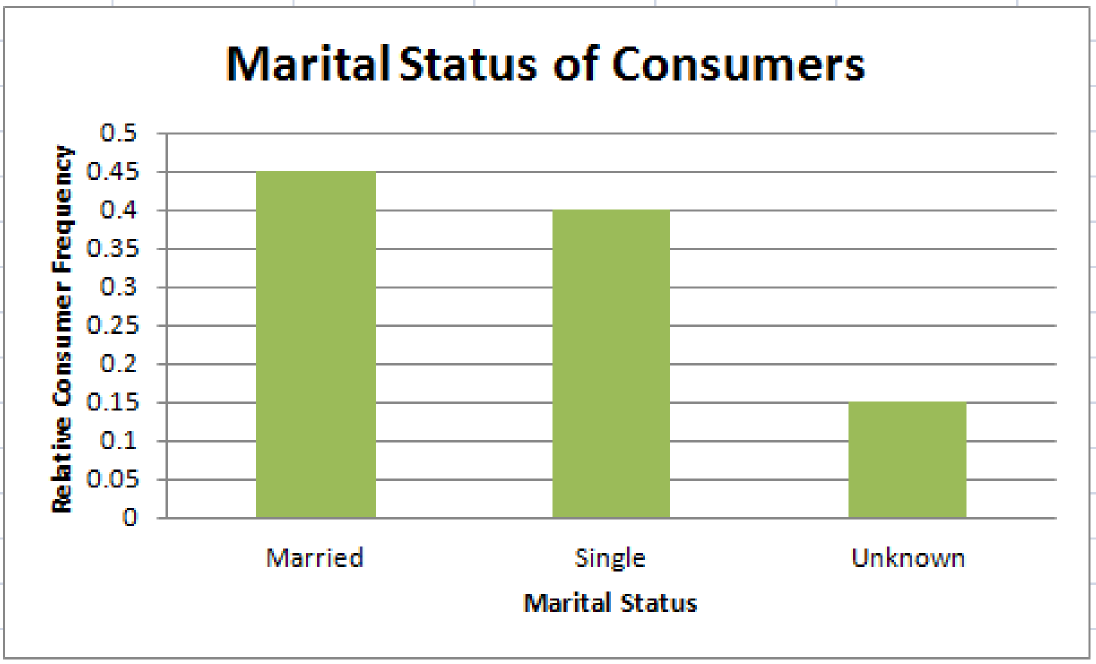


Figure 46: *A bar chart displaying the marital status of craft beer consumers.*

3.5.2.5 Religion

The two major religions of consumers are Protestant (51% of consumers) and Catholic (36% of consumers). Figure 47 summarizes the results in a histogram and a pie chart.

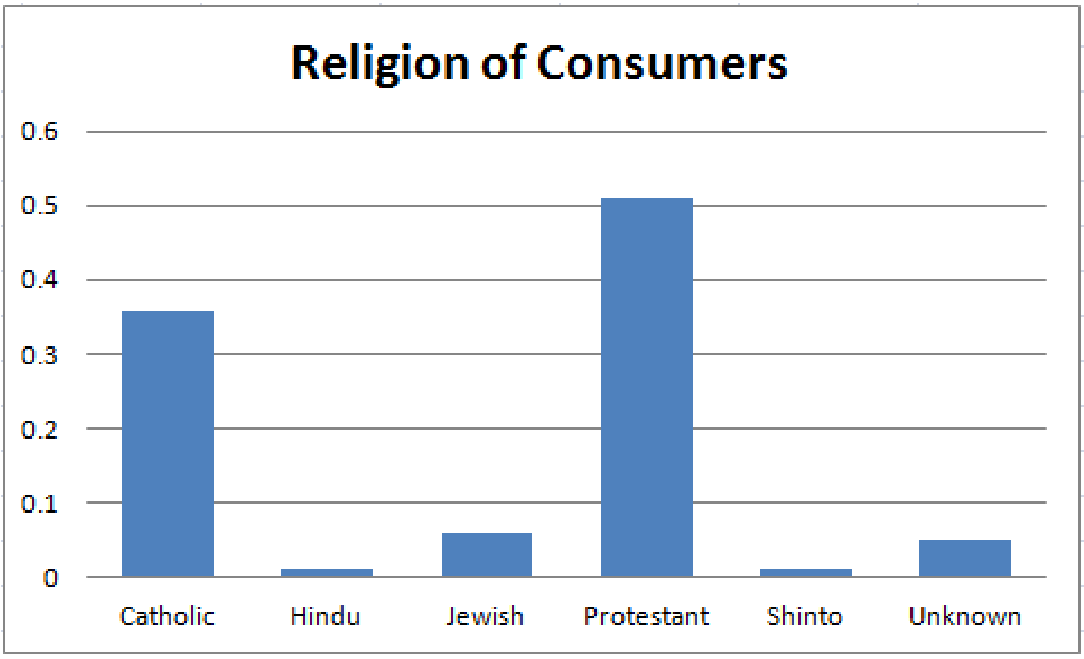


Figure 47: *A bar chart displaying the religious beliefs of craft beer consumers.*

3.5.2.6 Race

The predominant race of consumers is White (about 72% of consumers). The results are summarized in a pie chart (see Figure 48).

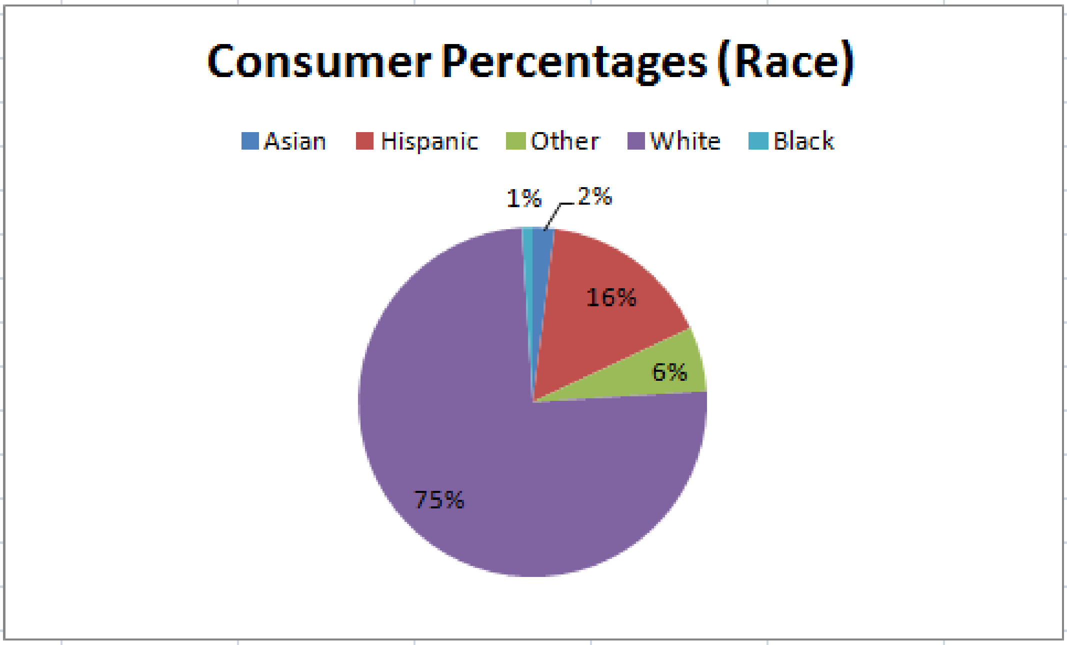


Figure 48: *Pie chart displaying the race proportions of craft beer consumers.*

4 Site Selection Model

4.1 Model Overview

For the development of the model a heuristic approach is chosen. The model is based on the geodemographic characteristics of the respective geographical areas. It incorporates the three demographic variables *gender*, *age* and *race*. The input parameters correspond to the number of people in a demographic group referring to persons that share the same demographic characteristics (e.g. White Males Age 25-34). There is a total number of 30 demographic groups which are shown in Table 5.The model corresponds to a geodemographic probability model which calculates the number of potential local craft beer consumers based on the empirically derived consumer characteristics of the previously conducted exploratory analysis.

4.2 Model Input – Geodemographic variables

The following geodemographic variables are incorporated in the model:

- Gender
    - Male
    - Female
- Age
    - 21 – 24
    - 25 – 34
    - 35 -44
    - 45 – 54
    - 55 +
- Race
    - White
    - Hispanic
    - Other

The three demographic variables are joined which leads to the 30 demographic groups which correspond to the model input parameters depicted in Table 5.

Demographic groups (Model Input Parameter)
White male 21-24
White male 25-34
White male 35-44
White male 45-54
White male 55+
White female 21-24
White female 25-34
White female 35-44
White female 45-54
White female 55+
Hispanic male 21-24
Hispanic male 25-34
Hispanic male 35-44
Hispanic male 45-54
Hispanic male 55+
Hispanic female 21-24
Hispanic female 25-34
Hispanic female 35-44
Hispanic female 45-54
Hispanic female 55+
Other Race male 21-24
Other Race male 25-34
Other Race male 35-44
Other Race male 45-54
Other Race male 55+
Other Race female 21-24
Other Race female 25-34
Other Race female 35-44
Other Race female 45-54
Other Race female 55+

Table 5: Demographic groups – model input parameters.

4.3 Model logic

The geodemographic model corresponds to a probability model based on the empirically derived results from the descriptive analysis of craft beer consumers.

The model calculates the expected value of potential customers within each geographical area. A geographical area usually comprises a large number of various demographics groups. The model is formalized as follows:

Notations:
C_p ... potential number of customers in geographical area (e.g. ZIP code area, catchment area etc.)
DG ... demographic group
m ... number of demographic groups in geographical area
POP_{DG} ... number of people in demographic group (model input → census data)
$P(DG_{Craft\ Beer\ Consumer})$ ... probability that a person of a specific demographic group is a craft beer consumer (→ derived from the exploratory analysis)

Model Formula:

$$C_p = \sum_{DG=1}^{m} POP_{DG} * P(DG_{Craft\ Beer\ Consumer})$$

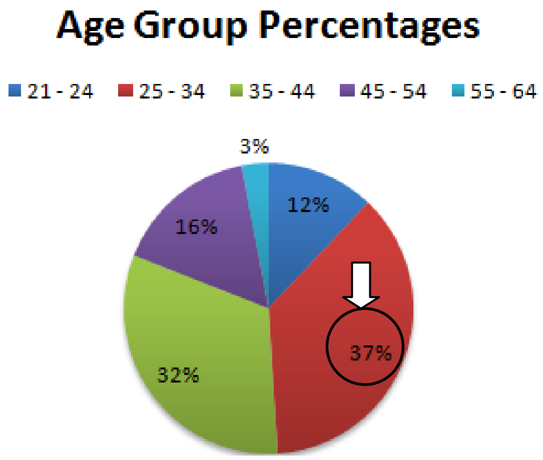
$P(DG_{Craft\ Beer\ Consumer})$ is based on the empirically derived statistics conducted in the previous section and is verified and compared with previously conducted external surveys such as the latest research of the company Mintel on craft beer consumption behaviors. Mintel has conducted a respective survey with the result that 53% of the average population drinks any kind of beer and that 23% of all beer drinkers favor craft beer. This leads to the statistic of 13% (approx.) of the overall population consuming craft beer. Based on the research of Mintel, assuming that the average of the total population consuming craft beer is 13% and based on the empirically derived statistics, Table 6 was calculated showing the probabilities that a person of a certain demographic group is a craft beer consumer:

Population group	$P(DG_{Craft\ Beer\ Consumer})$
White male 21-24	0.4260 (42.60 %)
White male 25-34	0.5520 (55.20 %)
White male 35-44	0.4619 (46.19 %)
White male 45-54	0.1987 (19.87 %)
White male 55+	0.0220 (2.20 %)
White female 21-24	0.1903 (19.03 %)
White female 25-34	0.2422 (24.22 %)
White female 35-44	0.2000 (20.00 %)
White female 45-54	0.0840 (8.40 %)
White female 55+	0.0080 (0.80 %)
Hispanic male 21-24	0.2929 (29.29 %)
Hispanic male 25-34	0.3764 (37.64 %)
Hispanic male 35-44	0.3852 (38.52 %)
Hispanic male 45-54	0.2629 (26.29 %)
Hispanic male 55+	0.0496 (4.96 %)
Hispanic female 21-24	0.1431 (14.31 %)
Hispanic female 25-34	0.1775 (17.75 %)
Hispanic female 35-44	0.1729 (17.29 %)
Hispanic female 45-54	0.1130 (11.30 %)
Hispanic female 55+	0.0176 (1.76 %)
Other Race male 21-24	0.1170 (11.7 %)
Other Race male 25-34	0.1580 (15.80 %)
Other Race male 35-44	0.1500 (15.00 %)
Other Race male 45-54	0.0840 (8.40 %)
Other Race male 55+	0.0140 (1.40%)
Other Race female 21-24	0.0530 (5.30 %)
Other Race female 25-34	0.0670 (6.70 %)
Other Race female 35-44	0.0600 (6.00 %)
Other Race female 45-54	0.0300 (3.00 %)
Other Race female 55+	0.0040 (0.04 %)

Table 6: *The probability that a person of a specific demographic group is a craft beer consumer.*

The calculation of the probabilities follows the following workflow (demonstrated by the demographic group White Male 25-34):

1. Taking 13% as a reference value of the overall US population consuming craft beer.
2. From the exploratory consumer analysis we know that 37% of craft beer consumers fall within the age group 25-34:



3. Calculate the proportion of the overall population that consumes craft beer and falls within the age group 25-34:

P(Craft Beer Consumers/25-34 out of overall population) *= 0.13 * 0.37 = 0.0481 = 4.81%*

4. Calculate the number of people out of the overall population that consume craft beer and fall within the age group 25-34:

Total Population US: *308745538*

Number of people 25-34 consuming craft beer out of overall population *= 308745538 * 0.0481 = 14850660 people of age group 25-34 consume craft beer*

5. Calculate the proportion of people in age group 25-34 that consume craft beer:

Population 25-34: *41063948*
Craft beer consumers in 25-34/overall POP: *14850660*

P(craft beer consumers in age group 25-34) *= P(25-34)*

= 14850660 / 41063948
= 0.3616 = 36.16 % of all 25-34 year olds consume craft beer

6. Calculate the proportion of people of age group 25-34 that are male craft beer consumers:

P(Craft beer consumers/male out of 25-34) *= 0.3616 * 0.7 = 0.2531*

7. Calculate the number of people out of the entire 25-34 group that consume craft beer, fall within the age group 25-34 and are male:

Number of male people 25-34 consuming craft beer out of all 25-34 *= 41063948 * 0.2531*

= 10395462 male people 25-34 consume craft beer

8. Calculate the proportion of male people in age group 25-34 that consume craft beer:

Population Male 25-34: *20632091*
Craft beer consumers in 25-34 AND Male / all 25-34: *10395462*

P(25-34/Male) *= 10395462 / 20632091 = 0.5038 = 50.38% of all male age 25-34 consume craft beer*

9. Calculate the proportion of male age 25-34 that are of white race and craft beer consumers

P(craft beer consumers/white out of males 25-34) *= 0.5038 * 0.75 = 0.3778*

10. Calculate the number of people out of males 25-34 that consume craft beer, fall within the age group 25-34, are male and are of white race:

Num white males 25-34 consuming craft beer *= 20632091 * 0.3778*

= 7796597 white male 25-34 consume craft beer

11. Calculate the proportion of white male people in age group 25-34 that consume craft beer:

Population White Male 25-34: *14123988*
Craft beer consumers in 25-34 AND Male AND White / all Males 25-34: *7796597*

P(White/25-34/Male) *= 7796597 / 14123988 = 0.5520 =* ***55.20% of all white male age 25-34 consume craft beer***

4.4 Model output

The number of potential customers per geographical area is calculated by applying the model formula which sums up the products of the number of people falling into a demographic group and the probability that a person out of this particular demographic group consumes craft beer:

$$C_p = \sum_{DG=1}^{m} POP_{DG} * P(DG_{Craft\ Beer\ Consumer})$$

$$\begin{aligned} C_p = \ & number\ of\ White\ Male\ 21 - 24 * 0.06 + number\ of\ White\ Male\ 25 - 34 * 0.19 \\ & + \cdots + number\ of\ Other\ Race\ Female > 55 \end{aligned}$$

5 Model Application Scenarios

5.1 Macro Level Analysis – Number of Potential Customers per ZIP Code

The objective of this analysis is to approximate the number of potential craft beer consumers for each ZIP code in San Diego County. Therefore the probability model developed in section 4 is run taking as input the demographic groups of each ZIP code. The output corresponds to an absolute score of potential customers and is also normalized by the overall population of the ZIP code to receive the percentage of craft beer consumers per ZIP code. The conceptual model is depicted in Figure 49.

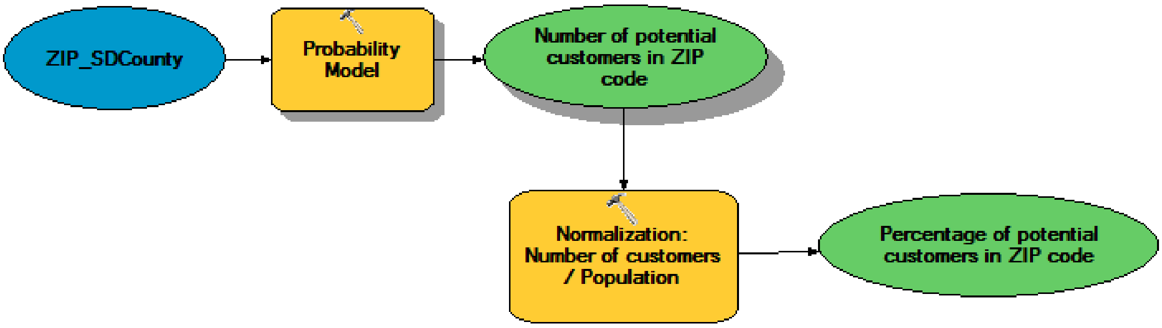

ZIP_SDCounty
Probability Model
Number of potential customers in ZIP code
Normalization: Number of customers / Population
Percentage of potential customers in ZIP code

Figure 49: *The conceptual model of the macro level location analysis scenario. The output corresponds to the potential number of craft beer consumers per ZIP code.*

A choropleth map is generated showing the output score of the model for each ZIP code and each area is colored according to the score. Figure 50 shows the number of potential customers per ZIP code whereas Figure 51 shows the percentage of craft beer consumers per ZIP code.

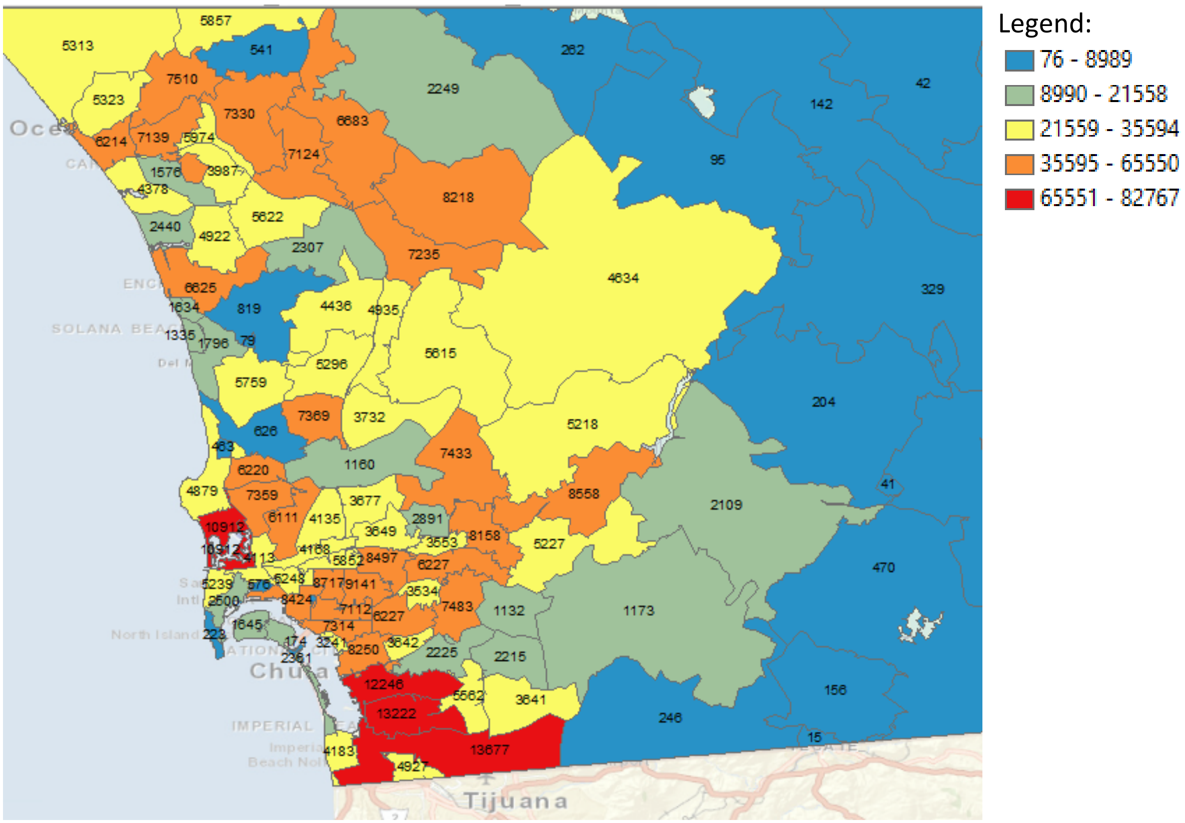


Figure 50: *Number of potential customers per ZIP code.*

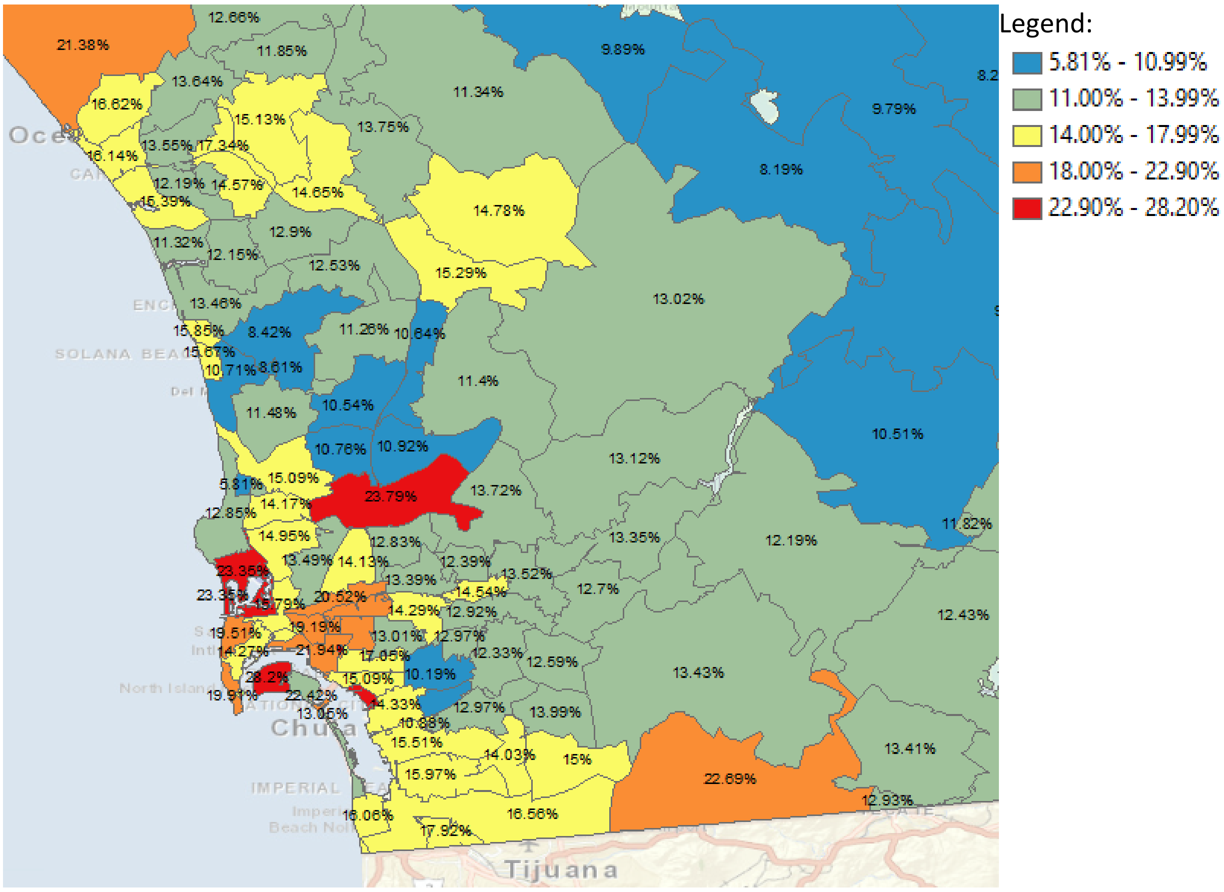


Figure 51: *Percentage of craft beer consumers per ZIP code.*

5.2 Micro Level Scenario – Site Suitability Analysis

This section describes a site suitability scenario with five candidate sites for sale or rent in San Diego County. The feasibility of each site is evaluated in respect of the people living within the catchment area of each site. First the general catchment area of microbreweries is approximated by examining the spatial distribution of frequently visiting customers of an existing microbrewery. Then a conceptual model is set up which outlines the workflow of this site selection scenario (see Figure 52). The input of the model is the point dataset of the candidate sites. A catchment area is generated around each site and the number of potential customers within each catchment area is calculated.

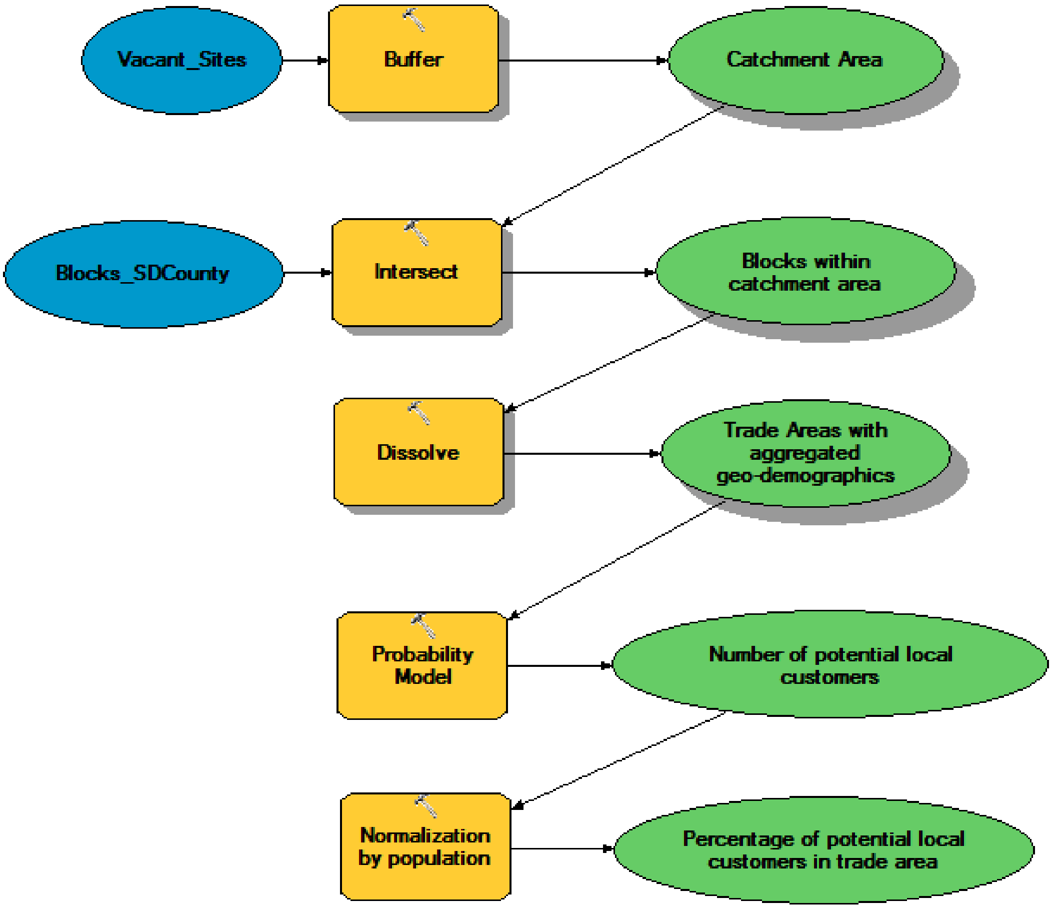

Vacant_Sites
Buffer
Catchment Area
Blocks_SDCounty
Intersect
Blocks within catchment area
Dissolve
Trade Areas with aggregated geo-demographics
Probability Model
Number of potential local customers
Normalization by population
Percentage of potential local customers in trade area

Figure 52: *The conceptual model of evaluating the feasibility of candidate sites.*

The conceptual model holds the following workflow:

1. Assess local catchment area (based on data of previous consumers of existing breweries)

2. Create a buffer representing the local catchment area

3. Intersect catchment area with census blocks, and extract blocks that fall within the catchment area.

4. Aggregate all blocks within a catchment area to one entity including the respective geodemographic attributes.

5. Run the probability model on each aggregated entity and receive as a score the **number of potential local customers per site**.

Catchment Area Assessment

Before defining the catchment area of the candidate sites, the general catchment area of local customers frequently visiting a microbrewery needs to be approximated. This estimation is based on the evaluation of an existing brewery in San Diego where the spatial distribution of frequently visiting local customers is examined (see Figure 53). The data is retrieved from the Twitter dataset.

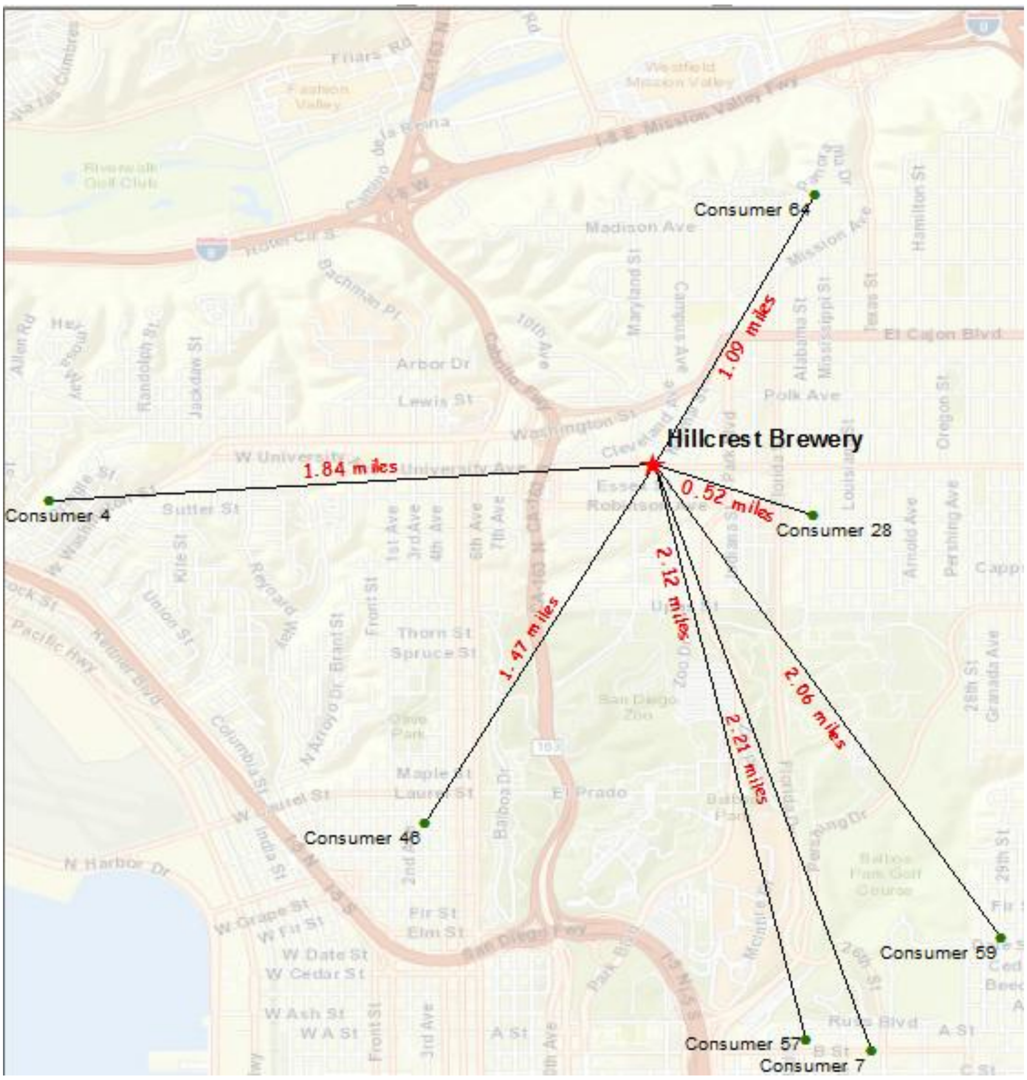


Figure 53: *Trade area assessment based on Hillcrest Brewery (red star) and a set of consumers frequently visiting the brewery (green dots). Additionally the Euclidian distances between each local customer and the brewery are depicted in red.*

The general catchment area of local customers is assessed with a 2.3 miles buffer (see Figure 54).

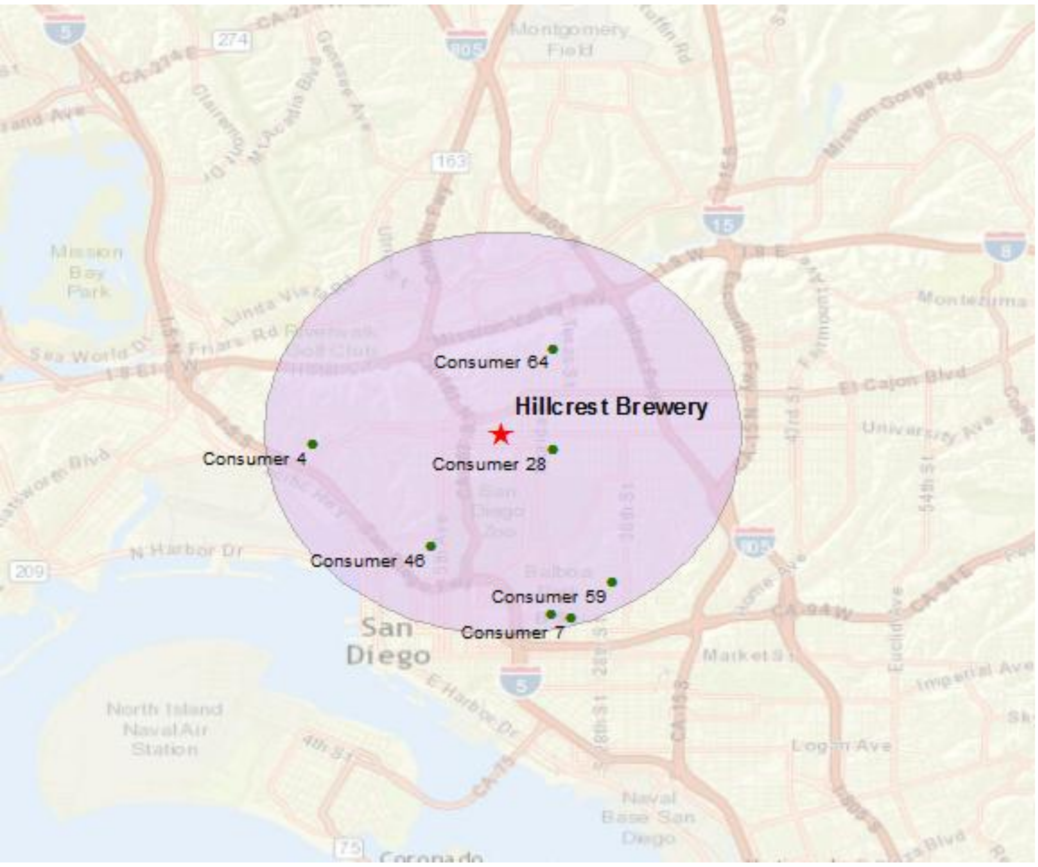


Figure 54: *Assessment of the general trade area of local customers of microbreweries with a 2.3 miles buffer.*

The conceptual model is implemented in ArcGIS and is run for each candidate site. The output score corresponding to the number of customers within the catchment area of each site is displayed in Figure 55.

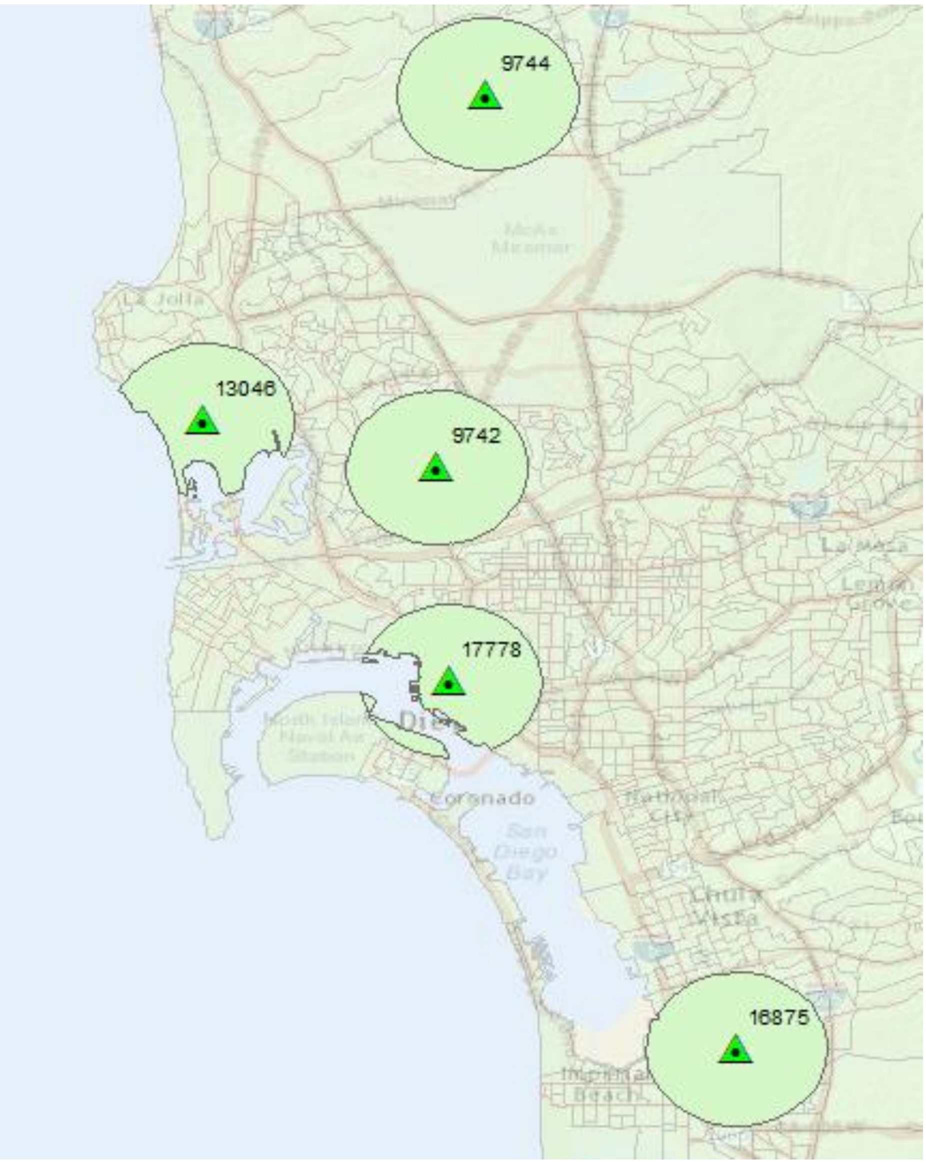


Figure 55: *The number of potential local customers for each candidate site (green triangles). The light green buffers around each site correspond to their respective local catchment areas.*

6 Conclusion and Future Work

This research consisted of two core parts. (1) An exploratory analysis of a Twitter dataset containing craft beer related tweets and (2) development of a site selection model for siting a microbrewery. The purpose of the exploratory analysis was to gain new knowledge about craft beer consumption behavior. This involved conducting a qualitative analysis aimed at examining the textual content of tweets to filter out messages which are unrelated to craft beer. Additionally craft beer ontology was built to allocate tweets to categories such as brewery names or types of beer. This ontology is based on keyword occurrences within the textual content of tweets and should serve as groundwork for future craft beer data analysis. A spatial analysis was conducted to get an understanding of the overall distribution of tweets in San Diego County. To this effect a raster based heat map implementing the Kernel Density Function was generated. The tweet densities for each neighborhood were calculated and visualized on a choropleth map. Additionally, tweets per brewery were computed and their numbers displayed on a point map. In a quantitative part of the analysis the demographic characteristics of the craft clientele were examined including variables such as *gender*, *age* or *race*.

The second part of this research comprised the design and development of a custom site selection model for new microbrewery opening scenarios. The objective was to estimate the number of potential customers within a specific geographical area. A heuristic approach was developed based on the knowledge of the demographic characteristics of craft beer consumers gained from the exploratory data analysis. The model follows the logic of a geodemographic probability model taking as an input the demographic characteristics of the geographical area to be examined and calculating the number of potential craft beer consumers within that area. Two model application scenarios were developed including a macro level analysis estimating the number of potential customers per ZIP code, for ZIP codes within San Diego County. A micro scale site suitability analysis was conducted to evaluate the feasibility of vacant sites considered as site options for a new microbrewery.

This research can serve as groundwork for future studies. The analysis of Twitter data can be extended to the examination of spatiotemporal movement patterns of craft beer consumers. This would be enabled by the availability of temporal data encoded within each tweet, where analysts can examine the trajectories of people to get an understanding of which other places they travel to prior or after visiting a microbrewery. This spatiotemporal analysis could for instance identify certain behaviors such as 'pub crawling' where people visit multiple breweries in a short period of time. For this research only tweets within the County of San Diego were taken into account. Further research could also incorporate tweets posted by tourists or out-of-state customers in order to approximate the overall number of potential customers per brewery. The ontology developed in this study can be expanded to additional categories, and automated systems can be implemented, which analyze craft beer related data quickly and efficiently. These systems could be based on an adapted version of the ontology developed in this research.

In the future, catchment areas of multiple existing breweries should be examined not only to assess the general primary catchment area of a microbrewery but also the secondary and tertiary catchment areas. An interesting study would be to model the number of customers actually visiting a certain microbrewery out of the pool of potential craft beer consumers living within the catchment area of the brewery. Therefore a calibrated version of the advanced huff model discussed in chapter 1.2.4 could be applied. This approach requires a detailed investigation of the model calibration parameters such as the question of how to quantify the attractiveness parameters of a microbrewery. A sample measure of attractiveness could be the number of existing microbreweries in the vicinity of a candidate site, which could have a positive effect on the number of customers due to the pub crawling behavior. The distances of how far craft beer visitors are willing to travel from brewery to brewery can be derived from a spatiotemporal analysis as described earlier. The availability of marketing-related information of existing breweries would enable the approximation of sales for a new microbrewery at a candidate site.

The microbrewery industry is expected to grow in the future and hence there is an opportunity for further research.

7 List of Figures

8 List of Tables

9 References

Literature

Applebaum W. (1968). "*The Analog Method for Estimating Potential Store Sales.*" In C. Kornblau, Guide to Store Location Research. Reading, Mass.: Addison-Wesley.

Adnan M., Lansley G., and Longley P.A. (2013). "*A geodemographic analysis of the ethnicity and identity of Twitter user in Greater London.*" Proceedings of the 21st Conference on GIS Research UK (GISRUK).

Birkin M., Clarke G. (1998). "*GIS, Geodemographics, and Spatial Modeling in the U.K. Financial Service Industry.*" Journal of Housing Research, Vo.9, No.1, pages 87-111.

Breheny M.J. (1988). "*Practical Methods of Retail Location Analysis: A Review.*" In N. Wrigley, Store Choice, Store Location and Market Analysis (pp. 39-86). London: Routledge.

Brown P. (1991). "*Exploring geodemographics."* In Handling Geographical Information (eds, Masser, I. and Blakemore, M.)." Longman, London, pp.221-58.

Brubaker B.T. (2001). "*Site Selection Criteria in Community Shopping Centers: Implications for Real Estate Developers.*" Brigham Young University, Department of Architecture, Master Thesis.

Bruckner R.W. (1998). "*Site Selection: New Advancements in Methods and Technology.*" New York: Chain Store Publishing Corp.

Boteler A. (2009). "*The Gourmet's Guide to Cooking with Beer*." Quarry Books. p. 15. ISBN 978-1-59253-486-9. Retrieved 21 July 2011.

Carey H.C. (1858). "*Principles of social science.*" Philadelphia: J.B. Lippincott

Chen X., Wong D. W., Yang C. (2013). "*Evaluating the geographical awareness of individuals: an exploratory analysis of twitter data*." Cartography and Geographical Information Science, Vo. 40, No.2, pages 103-115.

Church R.L. and Murray A.T. (2009). "Business Site Selection, Location Analysis, and GIS." John Wiley & Sons, Inc., Hoboken, NJ, USA.

Culnan M., McHugh P., & Zubillaga, J. (2010). "*How large U.S. companies can use twitter and other social media to gain business value*." MIS Quarterly Executive, 9(4), 243–259.

Fuchs G., Adrienko G., Adrienko N., and Jankowski P. (2013). "*Extracting Personal Behavioral Patterns from Geo-Referenced Tweets.*" Paper presented at the 16th AGILE Conference on Geographic Information Science, 14 – 17 May 2013, Leuven, Belgium.

Ghosh D. and Guha R. (2013). "*What are we tweeting about obesity? Mapping tweets with topic modeling and Geographic Information System.*" Cartography and Geographic Information Science, 40:2, 90-102.

Gonzalez-Benito O. and Gonzalez Benito J. (2005). "The role of geodemographic segmentation in retail location strategy." International Journal of Market Research, Vol. 47, Quarter 3.

Harris R. (2003). "*Population mapping by geodemographics and digital imagery.*" In Remotely Sensed Cities (ed., Mesev, V.), Taylor & Francis, London, pp.223-41.

He W., Zha S., and Li L. (2013). "*Social media competitive analysis and text mining: A case study in the pizza industry.*" International Journal of Information Management, Vo. 33, Issue 3, 464-472.

Huff, D.L. (1964). "Defining and estimating a trade area." Journal of Marketing 28: 34-38

Lapoint K. (2012). "Microbrewing in the US: An overview of the microbrewery industry and a business plan for future success." Honor Thesis. Paper 9.

Li H. and Yu L. (2005). "*A GIS-based site selection system for real estate projects.*" Construction Innovation 2005, 5: 231-241.

Li L., Goodchild M.F., and Xu B. (2013). "*Spatial, temporal and socioeconomic patterns in the use of Twitter and Flickr.*" Cartography and Geographic Information Science, 40:2, 61-77.

Lindqvist J., Cranshaw J., Wiese J., Hong J., Zimmerman J. (2011). "*I'm the mayor of my house: examining why people use foursquare - a social-driven location sharing application".* In Proc. CHI '11, pages 2409–2418. ACM.

Lynn K.L. and Jackson H.O. (1997). "*Geodemographics: An introduction for Textile and Clothing Researchers.*" Journal of Family and Consumer Sciences, Spring 1997, 89, 1, ProQuest Research Library, p. 30.

Kulshrestha J., F. Kootl, A. Nikravesh, K. P. Gummadi (2012). "*Geographic Dissection of the Twitter Network.*" In Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, Dublin, June 4–7, 202–209. Palo Alto, CA: The AAAI Press

Noulas A., S. Scellato, C. Mascolo, and M. Pontil. (2011). "*An Empirical Study of Geographic User Activity Patterns in Foursquare.*" In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, July 17–21.

Ringo L.G. (2009). "*Utilizing GIS-based Site Selection Analysis for Potential Customer Segmentation and Location Suitability Modeling To Determine a Suitable Location to Establish a Dunn Bros Coffee Franchise in the Twin Cities Metro, Minnesota.*" Department of Resource Analysis, Saint Mary's University of Minnesota, Minneapolis MN.

Sen A. and Smith T.E. (1995). "*Gravity Models of Spatial Interaction Behavior.*" Heidelberg: Springer-Verlag.

Sleight P. (1997). „*Targeting Customers: How to Use Geodemographic and Lifestyle Data in Your Business.*" NTC Publications, Henley-on-Thames

Snyder D.W. (2012). „*Feasibility Analysis of a Microbrewery*". Faculty of Agribusiness Department, California Polytechnic State University, Bachelor Thesis.

Takhteyev Y., A. Gruzd, and B. Wellman (2012). "*Geography of Twitter Networks.*" Social Networks, Special issue on Space and Networks 34 (1): 73–81.

Thrall G.I. (2002). "*Business geography and new real estate market analysis.*" Oxford, New York, Oxford University Press.

Tobler W. (1970). “*A computer movie.*” Economic Geography, 46, 234-40

Wei Q. (2001). “*Data envelopment analysis.*” Chinese Science Bulletin, August 2001, Vol.46, Issue 16, pp 1321-1332.

Xu C., Wong D.W., Yang C. (2013). “*Evaluating the ‘geographical awareness’ of individuals: an exploratory analysis of Twitter data.*” Cartography and Geographic Information Science, 40:2, 103-115.

Online Resources

ESRI (2014). “How Original Huff Model works:” http://resources.arcgis.com/en/help/main/10.2/index.html#/How_Original_Huff_Model_works/00mm0000004w000000/

Marriam-Webster (2004). “*Definition of Social Media*:” http://www.merriam-webster.com/dictionary/social%20media

TechTarget, accessed June 21, 2014, “*Definition of Twitter:*” http://whatis.techtarget.com/definition/Twitter

Twitter (2012). “Twitter turns six.” https://blog.twitter.com/2012/twitter-turns-six