MBA Thesis

Site selection for small retail stores using sustainable and location-driven indicators

Case study: Starbucks coffee shops in Los Angeles

Vadym Sokol & Kristijan Jordanov

Supervisor
Viroj Jienwatcharamongkhol

Karlskrona, Sweden
June 2020

This thesis is submitted to the Department of Industrial Economics at Blekinge Institute of Technology in partial fulfilment of the requirements for the Degree of Master of Science in Industrial Economics and Management. The thesis is awarded 15 ECTS credits.

The author(s) declare(s) that they have completed the thesis work independently. All external sources are cited and listed under the References section. The thesis work has not been submitted in the same or similar form to any other institution(s) as part of another examination or degree.

Author information:

Vadym Sokol
vs.vadymsokol@gmail.com

Kristijan Jordanov
kristijan.jordanov@gmail.com

Department of Industrial Economics
Blekinge Institute of Technology
SE-371 79 Karlskrona, Sweden

Website: www.bth.se
Telephone: +46 455 38 50 00
Fax: +46 455 38 50 57

Abstract

Site selection decisions remains a complex yet crucial process for strong business performance. Despite the extensive number of publications in this field, the emergence of new data collection technique, improved location analytics, and changes in consumers' preferences call for testing of new models and hypothesis. This study compares traditional site selection indicators (e.g. property size, proximities, competition, and demographic profiles) with novel site-selection indicators (e.g. environmental sustainability performance and socio-demographic characteristics from Tapestry data). By investigating a case study of Starbucks coffee stores in Los Angeles, we argue that environmental sustainability performance and socio-demographic Tapestry segments correlate with business performance indicators of small retail shops in two ways. First, higher sustainability scores result in increased foot traffic, and by extension increased business performance. Second, Tapestry segmentation stands as significant indicator of business performance in site selection modeling – specifically, by demonstrating the significant correlation between socio-demographic consumers' segments and the number of visitors per location. The output of this study offers an alternative location-driven site selection method, important for businesses and key industry-players in sharpening location-allocation decision-making processes.

Key words

Site selection; business performance; decision-making; location-driven decisions; modelling; small retail stores; sustainability; LEED; GIS; Tapestry segmentation; foot traffic; consumers; GLM; proximities; demographics; Starbucks; ESRI.

Acknowledgements

We would like to acknowledge ESRI Inc Business Analyst department for the opportunity to write this work under their guidance. Big thanks to UberMedia for providing foot traffic data and ESRI Inc for allowing the usage of their Business data and ArcGIS software that was crucial for this thesis. We would like to thank our tutor Viroj Jienwatcharamongkhol for the supervision and valuable guidance during the project time.

Stockholm, June 2020

Vadym Sokol & Kristijan Jordanov

Table of contents

List of Tables

List of Figures

List of abbreviations

AICc	Corrected Akaike Information Criterion (Model performance)
ESRI	Environmental Systems Research Institute
GIS	Geographic Information System
LEED	Leadership in Energy and Environmental Design
VIF	Variance Inflation Factor
GLR	Generalized linear regression model

1. Introduction

"Location, location, location!" This phrase is often used in the context of site location of real estate properties, following the first law of geography: "everything is related to everything else, but near things are more related than distant things" (Tobler 1970). The saying is still relevant, and it can be applied not just to real estate but also to business and infrastructure site selection. Site selection is pivotal in all sorts of businesses, including retail, service, wholesale, and manufacturing efforts. In fact, studies conducted by the Small Business Administration (SBA, 2019) and other organizations indicate that poor location remains one of the primary causes of business failure in America. Accordingly, strategic managers have always been struggling with spatial resource allocation decisions (Shelton 2016).

Site selection decisions are important in all types of contexts, from small retail shops to massive infrastructure projects. Whether it is a brand-new entrepreneur looking to establish their very first business site, a current business seeking to open another branch, or a large corporation planning the location of its new franchise, it is essential to take a fact-based approach. This means analyzing all the relevant data that are available to choose the best possible location. Farkas (2009) brings to light that decisions related to the location of facilities are usually complex, costly, and strongly impact the lifestyle of the surrounding community. Poorly selected sites can lead to lower revenues or even failure, whereas selecting a favorable geographic location can reduce costs and attract more customers (Steingold 2011). Valchou et al. (2016) also suggest that improved site selection would decrease the costs for decision-makers and could increase overall business performance.

In the retail sector, site location is of pivotal significance. Given the high level of competition in this sector, businesses must focus on attracting customers while also fighting their competitors. A good location can drive more customers to the shop while also placing it further away from competitors, leading to better revenues. Site selection can also help to reduce the costs of logistics and storage, which are among the major contributors to operational expenses in retail (Wang, Fan & Wang, 2018). Hence, in small retail businesses, such as coffee shops, location is tightly linked to key business indicators and, more importantly, survival.

Traditionally, site location modelling concentrates mainly on economic aspects. Nevertheless, global concern about climate change and the associated increasing awareness by customers of the importance of sustainable development demands that location decisions take into account not only economic but also environmental and social factors (Jang et al., 2015). Sustainability in selecting location, as Terouhid et al. (2012) suggests, is sought not only for environmentally harmful facilities (e.g. coal power plants) but for all business or facility types, including the retail industry.

Coffee shops are an excellent example of small retail businesses where location has a direct effect on foot traffic, sales and other performance indicators. As Han et al. explain (2018), the coffee shop industry is characterized by high competition and high concentration index, making it a very challenging market to navigate. For instance, there were approximately 20,000 coffee shops operating in 2011 in the USA (Jang et al. 2015). Additionally, profit margins in the food-service industry are generally low, and coffee shops may experience low net revenues due to the increased costs of operations (Kang, Lee & Kim 2010). The industry dynamic also plays a role here, with consumer tastes changing continually (Han et al. 2018). These characteristics of the coffee shop market make location critical to the success of both new and existing market players. A novel site selection model that takes into account the most updated socio-demographic factors, including sustainable and location driven parameters is pivotal for decision-makers in determining optimal placement for a coffee shop or relocation of existing one.

1.1. Problem discussion

The primary objective of any business is to maximize profit, and coffee shops business are no exception (Mankiw 2019). According to Grand View Research (2018), the size of the US hot drink industry was $17,356.6 million in 2017. The industry has enjoyed stable growth over the past years and is likely to grow further between 2020 and 2025 (Grand View Research 2018). Nevertheless, in order to sustain and grow business profitability, coffee shops must adjust to new trends, environments, and customer's needs, which means selling more environmentally friendly products and focusing on the sustainability of operations. A proper site location selection is a prerequisite for coffee shops to achieve these goals, specifically by increasing customer accessibility and reducing operational costs (Chen et al. 2012).

Even though site selection has been taken widely into consideration, it has been shown that past models or approaches for determining optimal site locations tend to quickly become obsolete due to changes in population patterns, commuting modes, consumption behaviors and so on (Aboulola 2018; Chen et al. 2012). A single demographic dataset can have over 2000 variation of demographical indicators (ESRI, 2020). This makes the construction of site selection models even more complex and time-consuming due simply to the number of variables that should be transformed and accounted for (Aboulola 2018). Consequently, it is of great importance that site selection models be frequently reanalyzed to examine updated factors and data involved in choosing location for retail stores and their relationship to site performance (Wang et al. 2014; Zentes et al. 2011).

Sustainability in the retail industry has attracted growing attention. Based on the Global Risk Report 2020, climate change is a major threat to the economy, with the issue having already increased environmental awareness among both businesses and consumers. The retail sector has seen significant changes over the past two decades. They have migrated to online platforms for everyday and luxury orders. Meanwhile, stores seeking to retain their brick-and-mortar presence are finding new ways to stand out (USGBC, 2019). In addition to impacts on the bottom line, retailers have also realized that "going green", or, in other words, giving weight to sustainability, offers them important customer advantages and shared values with new business partners. However, there is a lack of research studies examining how sustainability can be incorporated into site selection models applied in the retail sector. The above-mentioned issues suggest that examining the use of sustainable and location-driven indicators as socio-demographic segmentation in predicting coffee shop performance could help to improve site selection decisions in the small retail shops.

1.2. Problem formulation and purpose

As suggested in several studies by Aboula (2018), Amparo (2013), and Chen et al. (2018), it is important for companies to use novel site selection components as part of their site selection models in order to stay competitive in the small retail business. Therefore, site selection parameters must keep up to date with the most recent data collection techniques and the latest trends in technology.

Researchers have suggested potential benefits of using recent techniques as mobile activity and socio-demographic segmentation in predicting business performance (Aboulola 2018b; Grekousis & Thomas 2012). However, only a few studies have considered a location-driven segmentation of customer profiles as well as frequent capture of mobile activities so far. While population density has been proven to be a crucial factor for the business performance of retailers, mobile activity can provide insights about their exact positioning in time and space that could be applicable as a proxy of business performance. Meanwhile, socioeconomic, and demographic segmentation of population can explain the specific category of consumers, their social and economic status, as well as consumption preferences.

This study relies on location-driven, socioeconomic and demographic marketing segmentation of neighborhoods named Tapestry segmentation data, later in the thesis refer as "*Tapestry data*" or "*socio-demographic Tapestry segments"* provided by ESRI [1]

Although sustainability has become a prominent trend in business, there is a lack of data showing how sustainability indicators can be used as part of site selection models in the retail sector. Chen et al. (2012) describes sustainability as a progressive variable in the site selection process that should be tested more in future studies. Leadership in Energy and Environmental Design (LEED) standards could provide a way to predict future performances of retail businesses because they incorporate various sustainability metrics, such as water efficiency, energy and atmosphere utility, and general site sustainability (Son et al. 2012).

---

[1] Tapestry is a data product developed by Esri, an RPR partner and the leader in GIS software. Tapestry classifies U.S. residential neighborhoods into unique market segments based on socioeconomic and demographic characteristics.

This present study seeks to address this lacuna, by comparing traditional site selection indicators (e.g. property size, distance to stations, and demographic profiles) with novel site-selection indicators (e.g. environmental sustainability performance and socio-demographic Tapestry data). Looking to coffee retail stores, we argue that environmental sustainability performance and socio-demographic Tapestry data correlate with business performance indicators of small retail shops in two ways.

First, higher sustainability scores result in increased foot traffic, and by extension increased business performance. Second, tapestry segmentation stands as a significant indicator of business performance in site selection modeling – specifically, the presence of some socio-demographic segments significantly correlates with the number of visitors per location. For instance, we assume that with a higher concentration of pensioners near the store the business performance of that store would most likely decrease because the pensioners are not the typical consumer type of Starbucks products In this view, this study offers an alternative method of strategic site-location selection, crucial for businesses and key industry-players in sharpening corporate decision-making and improving profitability.

1.3. Delimitations

Although conducting interviews with coffee shop owners about the driving location selection components would have been valuable for this study, this was not possible due to time constraints and the lack of collaborating possibilities. Instead, this research draws on secondary data collection methods, as elaborated in section 3.1 (Saunders et al. 2016).

The second limitation relates to the inability using long-term analyses to measure business survival, for instance via IRR or NPV (MBA, 2019). The lack of financial and historical data for the selected sites have led to a cross-sectional study for the year 2019 (Saunders et al. 2016). Nevertheless, a longitudinal design is worth testing for future research.

The foot traffic data from mobile devices is used to measure business performance of the sites. Nevertheless, the data does not represent the actual counts of customers entered the stores, just the proxies of them. The data collection technique does not take into consideration stores located in airports, train stations, or shopping malls, given that customer motivation for patronizing these stores is likely to be incidental to the customers' main reason(s) for passing through such locations.

This study uses the LEED certification system as a sustainability ranking for the explanatory variable, which recognizes positive sustainable elements. Still, its limitation is that it does not penalize for implementing non-sustainable measures in the building design. The LEED scorecard gives a single point for positive achievement in each category and zero points for its absence. Since there is no method for deducing points, neutral and negative performance are equivalent. Taken together, the narrow perspective of sustainability based on positive scoring, as well as the failure to account for useful building life, undoubtedly leads to an inadequate measure of true sustainability (Denzer & Hedges 2011). However, with more than 12,500 retail spaces worldwide using LEED, it is clear that green building verification is an important driver of responsible development within the retail sector.

1.4. Thesis structure

The thesis includes six sections reporting on different components of the study. First, the Introduction and Literature Review will outline and discuss the problem of novel site selection decisions in the retail industry, assessed against the existing body of research in the field. Secondly, the Methodology section describes the research design and methods in greater detail, including data collection and analysis processes and operationalization. Next, the Data Analysis and Results sections present secondary data that was structured, filtered, and analyzed according to the theoretical problem framework and report on the results of the study. The Discussion section will explore the outcomes of data analysis, including their implications for the retail sector. The concluding section will establish the overall findings and conclusions of the research.

2. Theoretical framework *or* Literature review

2.1. Background

Given structural changes that occurred in the U.S. retail sector in the last few years, and the fact that real estate constitutes a substantial portion of retailers' assets and operating expenses, making accurate decisions surrounding store location has become pivotal to the strategy of most retailers. Due to their significance, decisions regarding retail site location must involve comparative assessments and analysis and take into account several alternatives and various factors affecting location choices (Timor, 2005). The literature on shop site selection is extensive, multi-disciplinary, and covers various approaches in operations, research, management science, economics, and marketing.

To prepare for the current study, a wide variety of articles has been analyzed. Studies by Onut et al. (2007), Valchou et al. (2016), and Aboulola (2018a) focus on the meta-analysis of various site selection techniques. Overall, most of the decision criteria are driven by economic factors, such as market trends, labor, raw materials, and transportation, as well as some non-economic factors, such as quality of life. Nevertheless, none of these studies reached the same conclusion. This means that site selection decisions are strongly dependent on the type of business, the place, and the criteria chosen as to where they will settle (Valchou et al. 2016).

To select the components relevant to coffee shop locations and business performance, a broader literature review was conducted, including various case studies in different industries, from small retail shops to teahouses allocation (Chen et al. 2018). Interesting findings were discovered about business performance indicators. Most articles used surveying and financial information to reflect business performance, although Chang et al. (2016) measured performance using social media reviews through custom applications.

Several studies focused on the importance of innovative combinations of statistical and decision-making models, as well as the usage of novel variables. For instance, Aboulola (2018b) look to the density of mobile activity as a novel variable to construct a site selection model for coffee shops. A study by Brunaer et al. (2010) has found that site selection techniques based on socio-economic neighborhood characteristics, such as rent prices, were effective in predicting favorable store locations and performance. Chen et al. (2012), in turn, focus on sustainability as a progressive variable in the site selection process that should be increasingly considered in the future model formations.

The aforementioned studies have analyzed a variation of demographic settings for site selection models. None of them however has applied the exact same set of variables due to the complexity of site selection models.

2.2. Economic theories of site selection

Even though business success (or failure) is dependent on a vast array of factors, site location is often regarded as the most unique and crucial (Chang & Li 2019; Fox et al. 2007; Li & Liu 2012). The influence of site selection on business performance is often enduring, inflexible, and costly given that locale choice is associated with long-term capital investment and any location-related change usually comes with significant expense (Wang et al. 2014). A poor site selection decision, as Li & Lieu (2012) stated, can hardly be mitigated regardless of the success of pricing or marketing strategies afterwards. A sound location choice, instead, will result in high customer visits, i.e. high foot traffic, potentially leading to higher sales volume, profits, and market share (Turhan 2013). In addition, the more strategic site selection should not only take into account consumer size within the area, but also consumer demographic - specifically their inclination to patronizing the business. Thus, businesses should select locations with both a large market of consumers as well as consumers' willingness to patronize their stores.

It is a well-known fact that the presence of popular chain brands such as Starbucks are markers of neighborhoods with high socio-economic levels (Carapetian 2017). This phenomenon is in line with the consumption theory of gentrification (Lees et al., 2010). Consumption theory posits that, as household income increases, so does household spending. Thus, people from better socio-economic backgrounds are more likely to actively shop and dine out, consequently driving local consumption up (Clower & Johnson 2017).

On the other hand, the term "sustainable consumption" has emerged. Van den Bergh, 1999 stats that sustainable consumption "reflects that most of the environmental and resource problems caused by humans are ultimately the result of consumption and life-styles". The research points out that there is a gap between the traditional economic models that are based on profit maximization and the increasing interest for the sustainable consumption.

Various studies have identified relations between site location and customer behavior, which in turn affects store performance. For example, the Law of Retail Gravitation, invented by Reilly (1931), demonstrates an inverse relationship between road distance and shopping frequency. Huff's probabilistic model developed in 1963 indicates that customers' choices were positively related to the size of shopping centers but negatively related to their distance from customers, and to the service/product quality of their competitors (Huff 1963). A large number of studies tried to expand Reilly's and Huff's models afterwards by including additional factors in order to gauge and adapt to changing shopping behaviors (e.g. Satani et al. 1998; Tong & Tong 2006). In general, past literature confirms a statistically significant relationship between site location and business performance, and continuing efforts have been made to determine emerging factors affecting this relation.

2.3. Traditional site selection indicators for retail

Economic theory defines that the main factors for site selection considers proximity to demand, suppliers, transportation infrastructure, labor, etc. (Bagchi- Sen & Hayter, 2001). The site selection has one key determinant – demand (i.e. when customers visit the location). Therefore, defining the attractive site location that would generate a high demand and would minimize the costs is crucially important.

The location-driven site selection provides insights about the business performance of the existing sites as well as valuable qualitative findings for future planning (Aboulola, P3, 2018). Therefore, it is necessary to consider relevant site selection components while structuring a statistical model that supports the decision-makers.

Traditional or most common components in constructing site selection models for small retails shops has been taken from the literature review (Baranzini and Ramirez, 2005; Aboulola 2018, Chen et al 2018).

1. *Property size*

Property size is one of the most common parameters to consider while analyzing business performance of any retail industry (Vend, 2018). Since coffee shops have small margins, it is crucial to maintain the law rate of revenue per square meter while satisfying “as many as can fit" customers with a cozy environment. In this view, property size directly correlates with consumer demand. As property size increases, so will consumer demand. Moreover, Starbuck, in particular, is known as a place for remote workers which require some extra personal space within the store. Therefore, coffee shops should make a trade-off decision by, on one hand, pleasing the customers with extra space while hopefully increasing sales and, on the other hand, trying to lower the revenue per square meters ration. It is validated by earlier studies that larger stores have higher net sales. This study tests whether it is relevant for our case study and whether it shows the same significant trend towards foot traffic.

2. *Number of stations*

Proximities have been mentioned in all site selection articles that we have reviewed. Most of the articles suggest that sites with closer distance to major infrastructure points as roads and parking slots are more suitable and would most likely attract higher demand.

This study is not the exception, it is assumed that the proximity to major infrastructure points would be a significant predictor for site selection of coffee shops. Hence, as we focus on sustainable decisions, only the walkability from major public transportation stations as bus stops and metro stations is considered. Meaning, that the shorter the walking-distance from a nearest bus station to a nearest store the higher the site attraction for the customer which leads to higher demands and higher sales. Instead of using traditional Euclidean distance, this study focuses on the exact geometry of the walking-distance buffer. This latest spatial network modelling functions in combination with latest mapping technology to calculate the amount of stations located within exact walking proximity from a site.

3. *Number of competitor stores surrounding the site*

The competitiveness of an environment is widely accepted as a factor driving store performance (Reinartz & Kumar 1999). The concept of spatial competition is described as proximity between stores that either provide the same products/services (direct competitors) or different ones (indirect competitors). Competition's influence on business performance has been considered among convenience stores, banks, grocery stores, and others (e.g. Lord and Wright 1981, Schmidt and Lee 1979). Since disposable income is inelastic, it was found that the more alternatives or competing options that exist nearby, the lower chance customers will visit or patronize a particular store (Kelly & Emlen 1993; Turhan 2013). Schmidt and Lee (1979) suggest that competitors often try to enhance their spatial monopoly by locating themselves as far from each other as possible to allow consumers to distinguish them based on distance. Accordingly, this study assumes that the number of competing stores has a negative impact on foot traffic of a certain Starbucks shop.

4. *Demographics*

Demographic attributes provide valuable insight for decision makers, especially with regards to making the site selection decisions. Common parameters like total population around the site gives an estimated number of potential customers. In addition, some demographic indicators such as age group can predict business performance better than others For instance, if we assume that the mean age for Starbucks customers are young and middle aged people, then a key determinant of site selection should reflect this age demographic

It is also worth considering what kinds of population live in the areas, differentiated by age, annual income, race, education, annual spending, and even specific consumption patterns. These categories all provide important cues for site-selection. For instance, it can be measured how much a certain population in a specific area spends, on average, on coffee?

Even though demographic data delivers a variety of valuable variables, it is time-consuming to account all the variables in the site selection model. There are over 2000 characteristics to consider and many of them are interdependent. Every variable can be differently correlated with specific industries and products and therefore should be statistically tested.

To determine which demographic variables can explain most accurately the business performance of Starbucks stores, this study tests, in new ways, demographic components already considered elsewhere (Aboulola, 2018): Total spending, total households, median age, education level, diversity index, spending on coffee, unemployment rate, crime index, business concentration.

2.4. Novel site selection indicators for retail

A variety of recent research studies such as Clark et al. (1997) indicated the need for novel indicators to be taken into account during location selection to address current problems in the field, including changes in the market environment and in consumer behavior. For Clark et al, early models remain weak considering the growing complexity of the decision-making processes. Therefore, novel approaches should be continually tested. This study suggests several novel site selection indicators for coffee shops such as:

1. Tapestry data (socio-economic, demographic segments of the population by ESRI)

As discussed in the theoretical framework, the presence such brands as Starbucks, marks neighborhoods with a good socio-economic environment (Carapetian 2017). Considering the literature review (Clower & Johnson 2017, Brunaer et al. 2010), and consumption theory in particular (Lees et al., 2010), it is possible to suggest that the use of socioeconomic, demographic segments of neighborhoods could also improve site selection model performance, as well as generate necessary insights for making better decisions about strategically-beneficial coffee shop locations.

This study draws on data surrounding the socioeconomic demographic segmentation of neighborhoods provided by ESRI Tapestry dataset[2] (ESRI, 2020). The significance of ESRI's Tapestry data lies in the fact that it applies modern data collection and modeling techniques to define socio-demographic clusters forming neighborhoods. As explained by Grekousis and Thomas (2012), Tapestry Segmentation utilizes census data, divergent residential socio-economic traits, and distinct shopping characteristics to identify and categorize consumers into 65 market segments, which are then further summarized into 14 LifeMode groups (Appendix, Table 8). For instance, Trendsetters, as a sub-group of LifeMode 3 (namely, Uptown Individuals), as its name suggests, consists of presentable singles who are fashion-oriented, tech-savvy, enjoy exploring life through travelling, and have not yet aimed for long-term settlement. Therefore, Trendsetters often do not possess houses or vehicles and tend to spend their disposable income on trendy and stylish entertainment-related activities and products. Since they are more often than not well educated and socially and environmentally responsible, this group often has high-paid jobs that allow them to live in upscale rental apartments and purchase green products as they are wont to do. Statistically, the Trendsetters sub-group has a median age of 36 and a median household yearly income of $63,100.

[2] Tapestry Segmentation is developed by combining traditional cluster analysis with state-of-the-art and robust data mining approaches. This combination allows Tapestry to take advantage of massive geodemographic databases whilst limiting the influence of outliers, making it efficient for analyzing small areas. Tapestry concentrates on input data, including household and housing characteristics, that demonstrate and drive the most distinguishable and unique spending behaviours. In 2010, Tapestry identified 65 market segments, which increased to 67 in 2019. These 67 segments can be grouped by similarities either in locale (5 Urbanisation groups), or in shared demographic figures (14 LifeMode groups).

With the analytical information as mentioned in the above example, Tapestry data could benefit site selection decisions by providing more detailed information about market segments residing in particular neighborhoods as well as indicated their preferences in consumption. This, in turn, could prevent businesses from opening stores in locations lacking customers from their target markets, thus increasing the odds of success. Still, there are no studies evaluating the use of Tapestry data in small retail businesses and its effect on business performance.

Based on current literature, the study will seek to evaluate the influence of using Tapestry data for suitable site location of coffee shops. Where Tapestry data is available, this research will employ ArcGIS Pro for software modeling.

2. Sustainability - Environmental performance (LEED)

An important measure of site selection performance are sustainability indicators, which, according to Chen et al. (2012), have grown significantly in terms of their role in the decision-making process and should thus be tested more in future studies. A report from the USGBC (2019) notes that, when retailers incorporate sustainability into their strategic decisions, they set themselves apart from their peers. Sustainability is crucial in the modern world for two main reasons. First, businesses that demonstrate their commitment to sustainability appeal to potential customers, who have become increasingly environmentally conscious over the past decade (Rendtorff 2019). Consequently, sustainability supports business profitability and competitiveness by distinguishing a business from its competitors. Location decisions are part of the company's sustainability strategy since they consider the use of natural resources, including soil, air, and water. This allows for hypothesizing in which the consideration of sustainability in site location decision-making will contribute to business image and improve performance. Secondly, sustainability is related to efficient resource use. Thus, taking sustainability into account while making location decisions helps to reduce future business costs (Rezaee 2016). As a result, firms considering location sustainability before deciding on the store site would be more likely to be profitable.

One particular issue that affects the use of sustainability as a component influencing location decisions is the low number of adequate site sustainability measures. The studies considered as part of the literature review did not offer any clear indicators of sustainability rankings in the literature that would consider the whole spectrum of environmental indicators, including energy efficiency, building material, green area indications, and others. For this reason, the study uses an approach that has not been tested before in retail site location selection. This approach involves measuring the sustainability of sites using LEED certification, or Leadership in Energy and Environmental Design. LEED is a program offered by the U.S. Green Building Council, which allows companies to quantify the sustainability of their buildings (Larson DG 2019).

LEED measures site design, water efficiency, indoor environmental quality, energy efficiency, and responsible material selection in building projects. Participation in the LEED process is voluntary, although it demonstrates social responsibility and environmental stewardship. 'LEED for Retail' is a version of LEED certification that is designed for stores, restaurants, and banks. Through LEED certification, building projects receive an internationally recognizable, third-party, verification that shows the world that the company takes sustainability seriously and integrates this concept into strategic decision-making (Larson DG 2019).

Regular changes in living conditions, consumption behavior, and increasing consumer awareness of environmental and social issues require location selection models to be frequently modified and updated. Crucially, these variables remain intimately connected with marketing initiatives (i.e. the way goods and services are market). As signaling theory suggests (Conelly et al. 2011) access to these forms of information will serve businesses in increasing consumer demand, and by extension business performance. A review of past studies however reveals a shortage of state-of-the-art models that take into account these factors. This study responds to this need by incorporating the most updated socio-demographic segmentation dataset (Tapestry) and sustainability certification system (LEED) into the development of a model determining optimal locations for Starbucks coffee shops.

2.5. Hypothesis development

This study analyses the problem of coffee shop site selection using location intelligence techniques. Decision-makers face significant challenges when choosing relevant components when comparing the performance of different site locations. Therefore, a site selection model with relevant and novel components that helps to identify the best site in terms of future business performance could benefit current owners of coffee shops and their future investors (Chen et al. 2018). The spatial component allows tracking dynamic parameters like demographics, population, urban environment, traffic, and others and is widely used to test the theoretical models (Marks et al., 1996).

In order to address the theoretical issues discussed in the literature review, this study aims to answer the following question:

Research question: *How do environmental sustainability performance, and newest socio-demographic tapestry data correlate with business performance indicators of small retail shops?*

Research hypotheses

Two hypotheses are developed in order to answer the research question. These hypotheses examine the relationship between (1) business performance with sustainability performance indicators and (2) business performance with Tapestry segmentation.

Previous studies have proven the importance of foot traffic as one of the most critical Key Performance Indicators (Baen 2000, Haugen et al. 2007). Foot traffic indicates the number of visits customers pay to a certain store, therefore, showing evidence of its customer patronage and associated potential for increasing sales and revenue (Abrishami et al. 2017). Accordingly, this study also uses foot traffic as a gauge for business performance.

From the literature review, it is clear that in order to stay competitive in the long run, companies have to operate and grow sustainably and that by practicing environmental sustainability, companies can improve their financial performance (San Martin et al. 2017; Taryn et al. 2019). Thus, this study proposes that the business performance of sites will be positively correlated with sustainability performance indicators. To be more specific, the better the sustainability performance indicators are, the higher foot traffic is. This can be justified by the fact that sustainability facilitates innovation, opens new markets, reduces business risks and enhances the brand reputation of retailers – all of which build resilience. And the retail sector is increasingly committing to making a critical impact in their markets by certifying their facilities (USGBC, 2019).

> **Hypothesis 1**. *Business performance of selected sites is positively correlated with sustainability performance indicators. Higher sustainability score results in higher foot traffic.*

By analyzing the past research in site selection for small retail stores, it has been concluded that several novel indicators, such as mobile data and socio-demographic data, explained the business performance of sites in the past. Therefore, this study will test the second hypothesis that, with improved data collection techniques allowing to collect the most recent data, the site selection model will contribute to more accurate prediction of business performance, leading to more appropriate selection of coffee shops' location and a higher number of visitors when the shops come into operation.

> **Hypothesis 2**. *Tapestry segmentation is a significant indicator of business performance in site selection modeling. The presence of certain Tapestry segments significantly correlates with the number of visitors per location.*

The test of the hypothesis using GLM statistical analysis is described in the analysis section.

2.6. Summary

The goal of the literature search was to provide a foundation of knowledge, define research areas of focus, and determine gaps in past research (Saunders et al. 2016). The literature review shows that integrating new components into location selection decisions could make them more effective in the modern retail context. For instance, mobile activity could predict the popularity of a location among customers, whereas Tapestry data could help to assess the presence of the target market in the area with a lesser number of variables. Sustainability of a site could also have an influence on performance due to its impact on business reputation and costs.

While research suggests that integrating these data in location selection decisions would be beneficial, there is little research evaluating the relationship between the use of these variables in site selection and the store's business performance. Hence, the primary research objective for this study is to fill the knowledge gap in the research by defining the most relevant, up-to-date parameters for site selection decisions that could be used by contemporary coffee shop companies. The secondary objective is to solve the problem of site selection decisions by introducing an alternative and innovative way of strategic selection of business sites using modern location data collection techniques.

The key parameters of the study were site selection methodology, business performance, innovative solutions, GIS technique use, and relation to sustainability. The focus of the research was on new trends and location technology methods in site selection that would improve strategic and operational decision-making processes. The study expects that the business performance of sites will be positively correlated with sustainability performance indicators, because sustainability creates innovation, opens new markets, reduces business risks and enhances brand reputation for retailers – all of which build resilience. The retail sector is increasingly committing to making a critical impact in their markets by certifying their facilities (USGBC, 2019).

The output of the study will provide valuable input for decision-makers, the industry, and the scientific community since it will help to update and enhance the models used in site selection decisions, thus ensuring that they respond to the current trends in the retail market.

3. Methodology

3.1. Research method

The goal of this part is to define the right set of data, methods and statistical models that would validate our hypothetical assumptions.

The inductive approach has been chosen to explore data from Starbucks locations to devise a model based on site's patterns. The model output could be applicable in predicting or defining the optimum location from a business performance perspective.

This study uses a design science research approach (DSR) to create and evaluate IT artifacts to solve identified organizational problems (Hevner et al 2004). Specifically, the quantitative mono method was used to define the nature of the scope, better understand the problem of sustainable site selection, and provide the contextual background for the further analysis. By constructing a statistical model to predict the spatial location of retail business, the method enables the facilitation and illustration new insights into alternative methods of decision-making for sustainable site selection, as well as provide the valuable information for the future research.

Case study

The decisions about site selection criteria are strongly dependent on the type of business, the place and the criteria chosen as to where they will settle (Valchou et al, 2016). Consequently, a case study method has been chosen since accurate information about the industry, geography and business objectives are fundamental for examining dependencies of the hypothesis and its relationship to the world (Ch5, Saunders et al., 2016).

A case study analyzing sites of Starbucks coffee shops has been chosen for this research due to several reasons: well-known brand with locations all over the world which makes it possible to relate for other markets; with high customer satisfaction index and one of the best in class when it comes to innovation (Tkaczyk, 2016). Starbucks is a profit-oriented business with environmentally conscious customers and management (Starbucks Report, 2018). Another important criterion was LEED-certification. After screening and filtering LEED-dataset for retail it has appeared that Starbucks coffee shops in Los Angeles, USA have the highest number of LEED-certified stores. Moreover, the mobile data that is used for foot traffic or customers concentration has been also available for this particular area. Therefore, the geographical choice of Los Angeles areas has been used for analysis. Crucially, this research uses the whole population of data.

3.2. Data Collection

Secondary data collection has been used for the study due to data availability and limited resources for primary data collection. An advantage of secondary data collection is shorter completion time, given the very limited timeframe that was given for this study. As trustworthy source, the biggest GIS-software and geo data provider in the market (ESRI, 2020), has offered a unique opportunity to use their software as well as business data which includes high resolution socio economic, demographic and financial information about the sites located in the USA. Behind ESRIs business data lies Infogroup with a history of 45 years of marketing research data collection. Data vendor makes over 25M verification calls to ensure the data quality (Inforgroup, 2020). Business data, market share, demographics and Tapestry segmentation were spatially extracted from ESRIs business data library.

As for the environmental data the US Green Building Council database for LEED Retail certifications (USGBC Projects, 2019) will be used. This dataset provides valuable information about sustainability ranking (Platinum - highest, Certified - lowest).

The Ubermedia data provider (Ubermedia, 2020) collects high resolution information about concentration of mobile phones usage and activity per certain geographic area. The data then aggregates and provides the insights of foot traffic count for a given area. The exact buffer around each Starbucks site has been calculated by data provider's experts. The data has been collected with an hourly resolution and summarized in the dataset.

3.3. Validity, Reliability and Ethical issues

"How do we know?" – This question is frequently asked in designing research methods to strengthen its quality (Ch5, Saunders et al., 2016). By using available dataset provided by ESRI, with business, demographical and socioeconomical information, this research method and analysis can be replicated in other geographic areas or markets in the USA, within also in different industry.

The study is thought to be reliable because the data collection technique and analytic procedures reproduces consistent findings, in line with works by various scientists in other, different contexts. The secondary data collected by a team of professionals from ESRI and UberMedia minimizes the risk of error and biases. Nevertheless, triangulation as well as a manual validation of store locations in the secondary data was used to confirm the validity/credibility of data, revealing thus the "reality" of it. The internal validity is established when research with high accuracy demonstrates a causal relationship between variables. This is especially relevant for quantitative studies (Ch5, Saunders et al., 2016). This research is considered valid when the intervention is shown statistically and leads to a consequential result (Ch5, Saunders et al., 2016), as statistically validated and described in the chapters Methods and Analysis. As for the external validity, testing the sample data from other states or countries is recommended, to verify the generalizability of the findings.

This project proceeds officially in collaboration with ESRI, who provides certified data and licenses for the research. Furthermore, the authors as external observers in assessment of sites allowed for an unbiased examination of the results since there are no professional relations between the authors and the analyzed industry, Starbucks. Finally, in conducting this research, provisions have been set out to ensure that this research does not violate any ethical code of conduct. Specifically, all data here have been acquired through mutual consent and agreement with respective stakeholders, including ESRI. In addition, we have ensured no fabrication or manipulation of the data that has been provided and analyzed.

3.4. Data preparation

The lack of control over quality of secondary data and gaps in some information has been resolved with several data collection methodology and engineering operations (Queiros & Gonzalez, 2019):

1. Spatial search or data availability by using ArcGIS Business Analyst Software
2. Data screening and evaluation, defining relevance for the study
3. Data evaluation, when was it collected, what purpose, collection methods, data quality
4. The locations of each store have been additionally validated with Esri Basemaps
5. Prepare and examine the dataset for the model

ESRI Business data library provides information about retail businesses such as Business Name, Address, Number of Employees, Property Size and Sales Volume. Of importance, property size has already been classified either 1 with 1500 – 2499 sq ft or 2 above 2500 sq. After selecting only stores that generated positive sales numbers in the year of 2019. It has been found 146 Starbucks locations in the Los Angeles area from which 25 were inside airports or train stations and were excluded from the dataset.

A validation analysis using Esri Basemaps shows that two Starbucks locations had a wrong address, so they were shifted to a correct place. Another two locations had duplicated locations and were also excluded from the final dataset. In total, 121 stores were used for the model.

LEED data has been separately filtered and visualized on the map. We rank the data that is based on the LEED-ranking system on 3 categories: 1 - certified stores, 2 - silver certificate, 3 - gold certificate. A geographical join function has been used then to join to each site location with LEED data (Map 1). There were in total 22 stores LEED-certified (18% of all sites) with 1 gold, 4 silver, 17 just certified. Due to the small presence categories gold and silver were combined into one named LEED-certified with distinction.

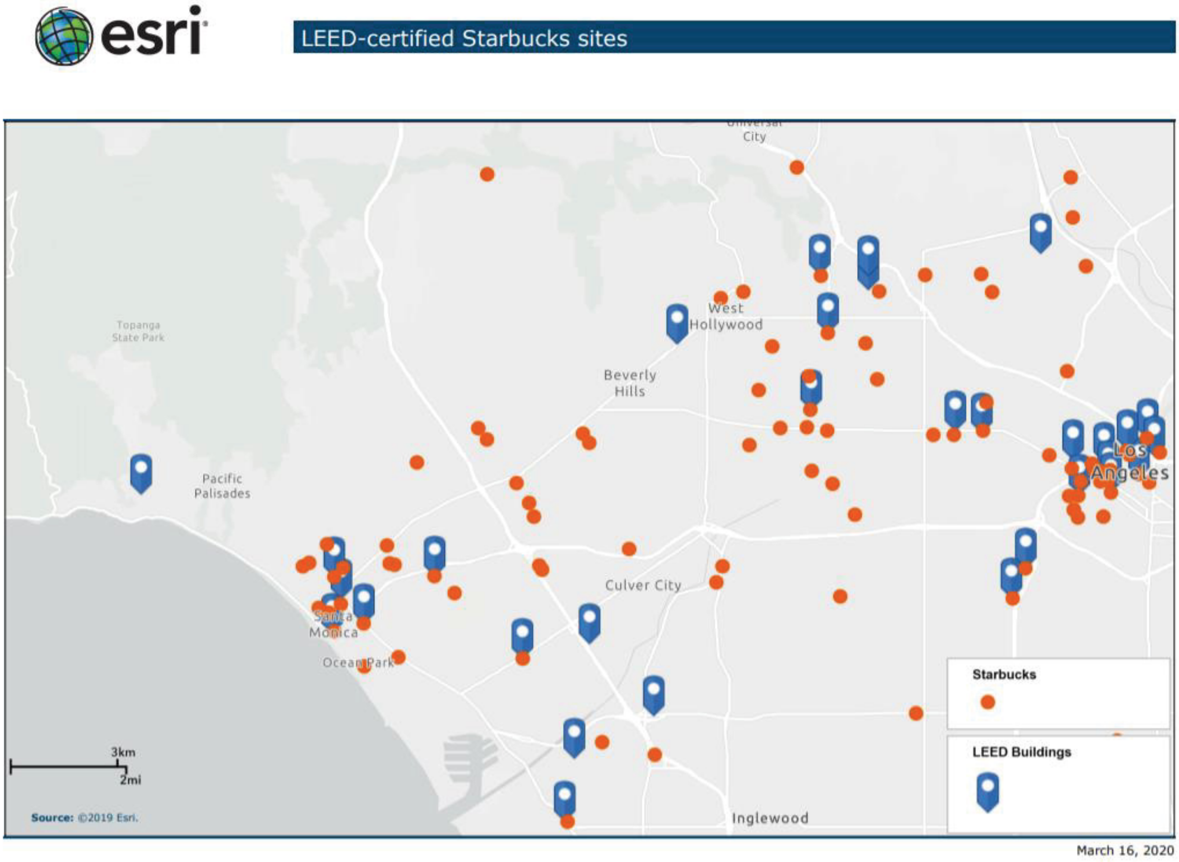


Figure 1. Starbucks stores with LEED-certified buildings

UberMedia data with foot traffic information has been extracted from a database to an excel file. Using the Pivot function, we calculated the annual count of foot traffic each site location. The geographical location was used to connect the foot trafficking data with UberMedia. All locations have been covered with provided data. In average each store had about 4058 activities with a standard deviation of 3697.

Data enrichment

Stores were drawn on the map using georeferencing. A service around each location was calculated with a walk-time buffer of five minutes. This buffer allows us to analyze the reachable area within a five-minute walk using high resolution road network and geographic network analyst functionality.

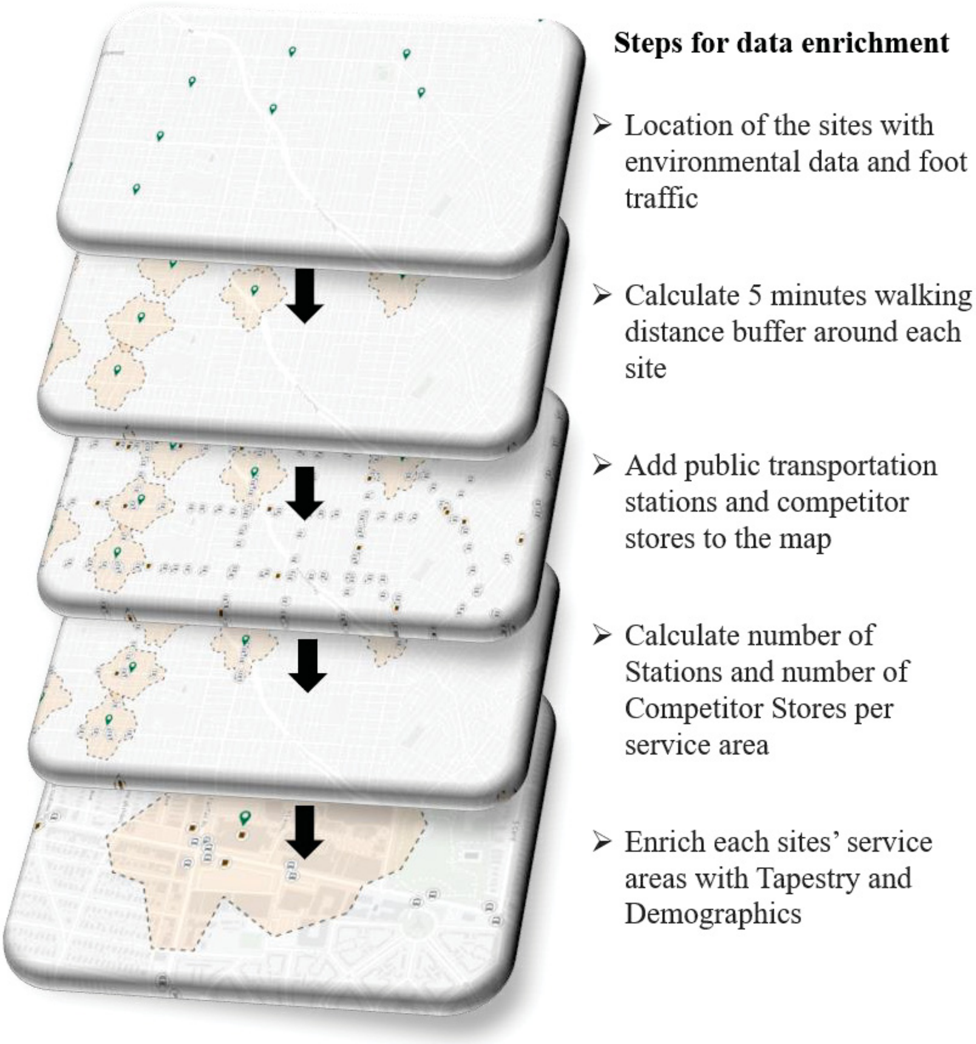


Figure 2. GIS-method for data enrichment

All the public bus, train, metro stations have been added to the map from open geo-library. In average it was found that there are about 14 stations per service area, with a standard deviation of 11. All coffee shops were considered to have the same market share with Starbucks and were therefore added spatially to the map as well. 53 of Starbucks stores had no competition at all. The average competition was 1.5 stores per site with a standard deviation of 1.9.

The service area around each site has been used to extract and count the amount of stations per store as well as number of competitors. Count of stations within 5 minutes walking distance represents more accurate information than the traditional Euclidean distance calculation approach because it takes into account exact route geometry.

Each service area (buffer area) has been overlaid with ESRI *demographics* as total population, age, race, level of education, household income and spending, business concentration, etc.

40 different demographic variables have been extracted from the Demographic dataset to enrich the sites. Data come from the latest release 2019. Variables with similar findings and that were not relevant for this site selection has been excluded from the analysis. The final set that has been used for the analysis: *Total Population, Income per capita, Median age, Education Bachelor, Diversity index, Spending on Coffee, Unemployment rate, Crime index, Business concentration (SIC).*

Table 1. Tapestry segments and Life Modes presented in the data

ID	Life Mode	Tapestry Segment	Count
3	Affluent Estates	Top Tier	3
11	Upscale Avenues	Urban Chic	6
11	Upscale Avenues	Pleasantville	4
11	Upscale Avenues	Pacific Heights	1
58	Uptown Individuals	Laptops and Lattes	11
58	Uptown Individuals	Metro Renters	18
58	Uptown Individuals	Trendsetters	29
2	Ethnic Enclaves	Southwestern Families	2
7	Middle Ground	City Lights	7
1	Middle Ground	Downtown Melting Pot	1
8	Senior Styles	Social Security Set	8
3	Midtown Singles	City Strivers	2
3	Midtown Singles	Young and Restless	1
2	Hometown	Family Foundations	2
20	Next Wave	International Marketplace	10
20	Next Wave	Las Casas	6
20	Next Wave	NeWest Residents	4
6	Scholars and Patriots	College Towns	2
6	Scholars and Patriots	Dorms to Diplomas	4

In Tapestry segmentation, there are in general 67 segments and 14 Life Modes (Appendix, Table 8). After spatial enrichment to our dataset, there were 11 different Life Modes and 19 Tapestry Segments presented in our sample (Table 1).

Each site's buffer then extracts information from the defined layers and summarizes the insights for each site in an attribute table (Appendix 1, Table 1).

Transform transformation and normalization

After the data processing (cleaning, formatting, combining) has been done we normalized and transformed attributes to prepare the variables for the regression model. The dataset has been normalized using a log-normal distribution (Table 2) which is a continuous probability distribution of a random variable whose logarithm is normally distributed, then calculated by:

Y= ln(X), where X is a default variable and Y would be a newly transformed normally distributed variable.

The categorical variables as size of the store size (Store size_1 - small; Store size_2 - large), LEEDs certification rankings (LEED_1 - certified; LEED_2 - certified with distinction), and Tapestry types (11 Life Modes and 19 Segments) has been converted to dummy variables (Table 2). Table 2 also shows that all the variables were checked on normal residuals distribution and were within the normal skewness ranges above +/-3 after normalization (Brown 2008). Table 3 show count and percentage of each Tapesty's Life Mode and Segment presented within the sample.

Table 2. Variables Normalization and Transformation

Variables	Mean / %	Std	Skew	Skew (log)	Transformation
FootTraffic	4,058.9	3,697.3	2.032	-0.363	Log
TotalPopulation	2,717.8	2,207.6	1.255	-1.891	Log
EducationBachelor	669.7	620.1	1.290	0.597	Log
IncomePerCapita	44,552.8	24,160.6	0.444	-0.738	Log
CrimeIndex	156.2	65.3	0.661	-0.143	Log
TotalBusiness	549.1	850.4	2.912	-0.382	Log
MedianAge	37.2	6.8	1.287	-0.329	Log
TotHouseholds	1,337.6	1,255.6	1.585	-1.581	Log
SpendingCoffee	162,792.4	152,836.8	1.477	-1.504	Log
Competition	1.4	1.9	1.624	1.382	Log
Stations	13.2	10.7	1.432	-0.847	Log
DiversityIndex	73.9	15.5	-1.182		Normalized
UnemploymentRate	5.6	3.6	1.775		Normalized
StoreSizeLarge	18%				Dummy (0, 1)
LEED1	14%				Dummy (0, 1)
LEED2	4%				Dummy (0, 1)
TapestrySegments					Dummy (0, 1)
TapestryLifeModes					Dummy (0, 1)

Table 3 Tapestry Segments and Life Modes

ID	Tapestry Segment	Count	% of sample	Tapestry Life Mode	Count	% of sample
1	Top Tier*	3	0.025	Affluent Estates*	3	0.025
6	Urban Chic	6	0.050	Upscale Avenues		
7	Pleasantville*	4	0.033	Upscale Avenues	11	0.091
8	Pacific Heights*	1	0.008	Upscale Avenues		
10	Laptops and Lattes	11	0.091	Uptown Individuals		
11	Metro Renters	18	0.149	Uptown Individuals	58	0.479
12	Trendsetters	29	0.240	Uptown Individuals		
32	Southwestern Families*	2	0.017	Ethnic Enclaves*	2	0.017
33	City Lights	7	0.058	Middle Ground	7	
36	Downtown Melting Pot*	1	0.008	Middle Ground	1	0.066
45	Social Security Set	8	0.066	Senior Styles	8	0.066
51	City Strivers*	2	0.017	Midtown Singles		
52	Young and Restless*	1	0.008	Midtown Singles*	3	0.025
56	Family Foundations*	2	0.017	Hometown*	2	0.017
60	International Marketplace	10	0.083	Next Wave		
61	Las Casas	6	0.050	Next Wave	20	0.165
62	NeWest Residents*	4	0.033	Next Wave		
66	College Towns*	2	0.017	Scholars and Patriots		
67	Dorms to Diplomas*	4	0.033	Scholars and Patriots	6	0.050

* Categories with less than 5% presence in the sample were excluded from the final models

3.5. Operationalization and Implementation

After variables were normalized (Table 2), we checked independent variables on multicollinearity. It has been found that in one demographic variable, *Total Population* had been used in calculating *Income per Capita, Education Bachelor*. Therefore, it has been excluded from the analysis to avoid issues with multicollinearity.

A dummy independent variables have been used for mutually exclusive categories as property size large, LEED-ranking and Tapestry categories (either present – 1, or absent – 0). A value of 0 causes that variable's coefficient to have no role in influencing the dependent variable, on the contrary, when the dummy is present, value 1, its coefficient acts to alter the intercept.

Dependent variable

Foot Traffic ($lnFT_i$) - is a dependent variable calculated by a logarithmically transformed number of annual counts of mobile activities per site and is used to measure the business performance per site *i*.

Independent variables

Traditional variables:

- *Property size large (DPS_i)* is a dummy variable which represents the presence of the large property sizes (above 2500 sq.ft) per site *i*.
- *Number of competitors ($lnNC_i$)* - a logarithmically transformed number of competitor stores per service areas of site *i.* Service areas around each site *i* have been calculated with 5 minutes walking-distance buffers.
- *Number of stations ($lnNS_i$)* - a logarithmically transformed summary of public transport stations of service areas of site *i.*
- Demographical variables have been spatially extracted and aggregated from each service area of site *i*. The most relevant of them were included in the model:

    i. *Income per Capita ($lnIC_i$)* – a logarithmically transformed total income per capita in USD of the population living in the service area around site *i.*
    ii. *Total Business ($lnTB_i$)* – a logarithmically transformed total amount of businesses per service area around site *i.*
    iii. *Education Bachelor ($lnEB_i$)* – a logarithmically transformed total amount of people with Bachelor's Degree living in the service area around site *i.*
    iv. *Crime Index ($lnCI_i$)* – a logarithmically transformed crime indexation per service area around site *i.*
    v. *Median Age ($lnMA_i$)* – a logarithmically transformed median age of the population per service area around site *i.*
    vi. *Spending on Coffee ($lnSC_i$)* – a logarithmically transformed total amount spent on coffee in USD per service area around site *i.*
    vii. *Total Households ($lnTH_i$)* – a logarithmically transformed total number of households in the service area around site *i.*
    viii. *Diversity Index (DI_i)* – a normalized diversity index, ranged from 1 to 100 per service area around site *i.*
    ix. *Unemployment rate (UR_i)* – a normalized unemployment rate, ranged from 1 to 100 per service area around site *i.*

Novel variables:

- *Tapestry Life Modes – presented as following categorical dummy variables*
    i. Upscale Avenues ($DLM1_i$)
    ii. Uptown Individuals ($DLM2_i$)
    iii. Middle Ground ($DLM3_i$)
    iv. Senior Styles ($DLM4_i$)
    v. Next Wave ($DLM5_i$)
    vi. Scholars and Patriots ($DLM6_i$)

- *Tapestry Segments – presented as following categorical dummy variables*
    i. *Urban Chic* ($DS1_i$)
    ii. *Laptops and Lattes* ($DS2_i$)
    iii. *Metro Renters* ($DS3_i$)
    iv. *Trendsetters* ($DS4_i$)
    v. *City Lights* ($DS5_i$)
    vi. *Social Security Set* ($DS6_i$)
    vii. *International Marketplace* ($DS7_i$)
    viii. *Las Casas* ($DS8_i$)

- *Environmental sustainability performance* (LEED-ratings) were presented as following dummy variables:
    i. *LEED1* ($DL1_i$) – LEED-certified buildings per site *i*
    ii. *LEED2* ($DL2_i$) - LEED-certified buildings with distinction per site *i*

3.6. Modeling

Enter method was used for regression models to ensure that all significant independent variables were directly entered into the equation. The linear regression model provides indications of which set of out variables or key indicators better explains the business performance of the site.

In order to make better decisions in the site selection process for small retail stores two groups of statistical models have been tested. The first one, "Traditional", is more commonly used in previous studies and was used as a baseline for our novel model. The second model called "Novel", enriched with alternative indicators. As Tapestry data presented in Life Modes categories with Segments subcategory, we ran two separate models: one with Tapestry Life Modes and another with Tapestry Segments. We intended to find which one of two Tapestry models performed better to further compare it with the "Traditional" model.

As Tapestry categories are formed using a set of demographic variables, it has been decided not to mix demographic variables and Tapestry data in the same model.

1. Traditional Model:

In the traditional model the business performance of a site is explained by the combination of traditional variables including property size, number of competitors and number of stations around each site as well as a set of demographical variables.

$$(1)\ lnFT_i = \beta_0 + \beta_1 DPS_i + \beta_2 lnNC_i + \beta_3 \ln NS_i + \beta_4 lnIC_i + \beta_5 lnTB_i + \beta_6 lnEB_i + \beta_7 lnCI_i + \beta_8 lnMA_i + \beta_9 lnSC_i + \beta_{10} lnTH_i + \beta_{11} DI_i + \beta_{12} UR_i + \epsilon_i$$

,where *Foot Traffic* ($lnFT_i$) *is* the dependent variable. β is the estimated coefficient of each variable. Dummy *Property size large* (DPS_i), *Number of competitors* ($lnNC_i$), *Number of stations* ($lnNS_i$), *Income per Capita* ($lnIC_i$), *Total Business* ($lnTB_i$), *Education Bachelor*

($lnEB_i$), Crime Index ($lnCI_i$), Median Age ($lnMA_i$), Spending on Coffee ($lnSC_i$), Total Households ($lnTH_i$), Diversity Index (DI_i), Unemployment rate (UR_i) are the independent. ϵ_i - *error term* that indicates the uncertainty in the model.

2. Novel Model:

2a) Novel model A examines the business performance (foot traffic count) of a site by the combining some traditional variables including property size, number of competitors and number of stations around each site as well as novel variables including categories of LEED-certified stores and groups of Tapestry Life Modes.

$$(2)\ lnFT_i = \beta_0 + \beta_1 DPS_i + \beta_2 lnNC_i + \beta_3 \ln NS_i + \beta_4 L1_i + \beta_5 L2_i + \beta_6 DLM1_i + \beta_7 DLM2_i + \beta_8 DLM3_i + \beta_9 DLM4_i + \beta_{10} DLM5_i + \beta_{11} DLM6 + \epsilon_i$$

,where *Foot Traffic ($lnFT_i$) is* the dependent variable. β is the estimated coefficient of each variable. Dummy *Property size large (DPS_i), Number of competitors ($lnNC_i$), Number of stations ($lnNS_i$),* dummy *LEED1 ($DL1_i$),* dummy *LEED2 ($DL2_i$)* and dummy *Tapestry Life Modes ($DLM1_i$* to *$DLM6_i$)* are the independent variables. ϵ_i - *error term* that indicates the uncertainty in the model.

2b) Novel model B analyses the relationship between the business performance and property size, number of competitors, number of stations, LEED-certified stores and groups of Tapestry Segments instead of Tapestry Life Modes presented in Model 2a.

$$(3)\ lnFT_i = \beta_0 + \beta_1 DPS_i + \beta_2 lnNC_i + \beta_3 \ln NS_i + \beta_4 L1_i + \beta_5 L2_i + \beta_6 DS1_i + \beta_7 DS2_i + \beta_8 DS3_i + \beta_9 DS4_i + \beta_{10} DS5_i + \beta_{11} DS6 + \beta_{12} DS7 + \beta_{13} DS8 + \epsilon_i$$

,where *Foot Traffic ($lnFT_i$) is* the dependent variable. β is the estimated coefficient of each variable. Dummy *Property size large (DPS_i), Number of competitors ($lnNC_i$), Number of stations ($lnNS_i$),* dummy *LEED1 ($DL1_i$),* dummy *LEED2 ($DL2_i$)* and dummy *Tapestry Segments ($DS1_i$* to *$DS8_i$)* are the independent variables. ϵ_i - *error term* that indicates the uncertainty in the model.

Model evaluation

The significance of the variables was measured using robust p-value (probability). P-value < 0.05 means that for 95% confidence level a p-value smaller than 0.05 indicates statistically significant relation. The robust probability is used in the result section to assess the variables significance to the dependent variable. The probability provides two major insights. First, we were able to know whether our hypotheses were false or true by analyzing the correlation between dependent and independent variables within the model contact. Second, we were able to define which variables to use for constructing a significant regression model in the future.

A variables coefficient sign points on the relationship between dependent and independent variables. A positive coefficient means that the linear relationship between variables is positive. The negative sign defines a negative relation between variables, meaning that with the decrease of the independent variables the dependent variable increases its values. The models also were checked upon standard residuals and multicollinearity to test whether the model was biased or not. Primarily Large Variance Inflation Factor (VIF) was used to indicate redundancy among explanatory variables. VIF values > 7.5 indicate redundancy among explanatory variables and means that variables with high VIF should be gradually removed from the model.

The statistical regression model analyses which variables are significant for predicting the dependent variable and which are not. The evaluation of the models is statistically represented by Multiple R^2 and Adjusted R^2 values. Higher multiple R^2 and adjusted R^2 values means better model performance. Another model performance indicator is Corrected Akaike Information Criterion (AICc) which was used to compare different models. The model that shows smaller AICc value is the better performed model.

Joint F-Statistic was used to measure an overall model statistical significance. The Joint F-Statistic method is based on p-value significance, where (p-value<0.05) means that the model is significant.

4. Empirical findings & Results

4.1. Model 1 - Traditional

Model 1a, with all traditional variables as store size, demographics, number of stations and competitors, run with all transformed demographic variables (Appendix Model 1a). Two variables *Total Households* and *Spending on Coffee* had very high VIF numbers, 106 and 77 respectively, and had to be gradually removed from the final model. VIF values above 7.5 should be removed from the models without causing any jeopardizing of it.

Model 1b with a final set of **traditional** variables as store size, number of stations and competitors, and demographics (with VIF < 7.5)

Dependent variable: *FootTraffic*. Independent variables: Demographics (*IncomePerCapita, TotalBusiness, EducationBachelor, TotalCrime, MedianAge, DiversityIndex, UnemploymentRate*), *Competition, Stations, PropertySize_Dummy.*

Model performance:

Variable	Coefficient	Robust_StdEr	Robust_Prob	VIF
Intercept	**12.613***	2.617	0.000	----------
IncomePerCapita	**-0.611***	0.228	0.009	3.575
TotalBusiness	**-0.209***	0.086	0.017	2.988
EducationBachelor	-0.027	0.099	0.785	7.243
TotalCrime	0.360	0.239	0.134	1.932
MedianAge	0.762	0.499	0.129	5.965
DiversityIndex	-0.013	0.007	0.078	2.132
UnemploymentRate	-0.045	0.023	0.050	1.769
Competition	-0.121	0.128	0.345	1.706
Stations	**0.278***	0.116	0.018	2.179
PropertySize_Dummy	**0.598***	0.209	0.005	1.108

N: 121; Multiple R-Squared: 0.295; Adjusted R-Squared: 0.231; AICc: 303.646; Joint F-Statistic 0.000*

Model 1b -Traditional. GLR model output

A model with traditional indicators has shown a model performance with **adj. R^2 0.231, R^2 of 0.295** (Model 1). Demographic variables as *IncomePerCapita* (p-value 0.008), *TotalBusiness* (p-value 0.016) were the most significant demographic variables with negative correlation. Followed by the unemployment rate (p-value 0.050) which was marginally significant with a negative trend towards foot traffic. *PropertySizeDummy* or store size above 2500 sq.ft. (p-value 0.005), amount of stations within 5 min walking proximities (p-value 0.018) were positively and significantly correlated with business performance indicator (foot traffic).

4.2. Model 2 - Novel

In addition to default variables as store size, count of stations and competitors we added two novel parameters – LEED-rankings and Tapestry Life Modes and Tapestry Segments. Due to the fact that socio-demographic Tapestry data includes in its calculations a combination of demographics variables (as age, income, etc), all the demographic variables have been excluded from the models to avoid multicollinearity. Firstly, we model the Tapestry Life Modes since it groups Tapestry Segments into specific modes. Secondly, we run the same model with Tapestry Segments.

Model 2a - Novel Model with Tapestry Life Modes and LEED

Dependent variables: *FootTraffic.* Independent variable: *Competition, Stations, PropertySize_Dummy, LEED1_Dummy, LEED2_Dummy,* Tapestry Life Modes (*UpscaleAvenues_Dummy, UptownIndividuals_Dummy, MiddleGround_Dummy, SeniorStyles_Dummy, NextWave_Dummy, ScholarsAndPatriots_Dummy*).

The model performance showed adj. R^2 0.211, R^2 0.280 (Appendix Model 2a). Since none of the Tapestry Life Modes showed significant relationship with foot traffic, the model has been excluded from the results. We focused our further analysis on the more detailed Model presented with Tapestry Segments instead.

Model 2c - Novel Model with Tapestry Segments and LEED

Variables with less than 5% presence in the sample were excluded from the final model 2c. Model 2b with all the Tapestry Segments presented in the Appendix (Model 2b)

Dependent: *FootTraffic.* Independent variables: *Competition, Stations, PropertySize_Dummy, LEED1_Dummy, LEED2_Dummy,* Tapestry Segments (*UrbanChic_Dummy, Laptops&Lattes_Dummy, MetroRenters_Dummy, Trendsetters_Dummy, CityLights_Dummy, SocialSecuritySet_Dummy, InternationalMarketplace_Dummy, LasCasas_Dummy).*

Model performance:

Variable	Coefficient	Robust_StdEr	Robust_Prob	VIF
Intercept	6.752	0.282	0.000	--------
Competition	**-0.274***	0.125	0.030	1.309
Stations	**0.069***	0.096	0.000	1.839
PropertySize_Dummy	**0.665***	0.217	0.003	1.169
LEED1_Dummy	0.069	0.229	0.763	1.121
LEED2_Dummy	0.483	0.424	0.257	1.142
UrbanChic_Dummy	0.062	0.388	0.388	1.446
Laptops&Lattes_Dummy	0.081	0.251	0.747	1.430
MetroRenters_Dummy	-0.105	0.212	0.621	1.472
Trendsetters_Dummy	**-0.425***	0.205	0.041	1.649
CityLights_Dummy	-0.124	0.420	0.768	1.286
SocialSecuritySet_Dummy	**-0.888***	0.309	0.005	1.493
InternationalMarketplace_Dummy	**0.570***	0.243	0.021	1.331
LasCasas_Dummy	0.479	0.379	0.379	1.221

N: 121; Multiple R-Squared: 0.296; Adjusted R-Squared: 0.211; AICc: 311.087; Joint F-Statistic **0.000***

Model 2c - Novel Tapestry Segments + LEED. GLR model output

Model 2 has performed similarly to Model 2a with **adj. R^2 0.211, R^2 0.296** (Model 2b). *PropertySize_Dummy* (p-value 0.002) and proximities indicators as number of *Stations* (p-value 0.001) and *Competition* (p-value 0.028) were among the significant independent variables. The LEED data has not shown any significant effect on the business performance, with a slightly upwards pointing curve. LEED certified stores *LEED1_Dummy* (p-value 0.762) were not helping the model. Although, and LEED-certified stores with distinction *LEED2_Dummy* (p-value 0.256) showed a better performance, the p-value was still statistically not significant. Tapestry Segments as *SocialSecuritySet_Dummy* (p-value of 0.005). and *Trendsetters_Dummy* (p-value of 0.041) showed negative significant correlation with the amount of foot traffic (business performance). On the other hand, segment *InternationalMarketplace_Dummy* with p-value 0.020 showed statistically significant positive relationship to business performance of the stores.

The explanation of the model performance and discussion of the results are provided in the next chapter Analysis & Discussion.

5. Analysis & Discussion

This section is focused on the analysis of the results. We describe in detail the findings of the study taking into consideration the research scope and the results of the model.

The purpose of this thesis was to investigate the effectiveness of the novel indicators on location-driven site selection modelling for Starbucks coffee shops. Several studies indicated that business success or failure depended on site selection decisions (Chang & Li 2019; Fox et al. 2007; Li & Liu 2012) which are complex and need to be constantly improved due to changes in the environment, infrastructure, competition and customers' behaviors.

By assessing business performance of Starbucks stores using foot traffic, we were able to test whether such indicators as environmental sustainability and socio-demographic segmentation can explain the foot traffic and whether they should be included in site selection models. In order to answer our research question, two hypotheses have been tested by using regression models. In addition, we tested the latest location driven GIS-technique to calculate service areas around stores to measure proximities from a sustainable point of view.

5.1. Traditional vs Novel Site Selection Model

We constructed the Traditional model based on the literature review to assess business performance by using a set of variables as store size, competition, distance to stations and various demographic parameters.

The significant model performance (adj. R^2 0.231, R^2 0.295) provides two major insights. First, it shows that our set of traditional parameters has been chosen correctly for this study. In addition, it shows that foot traffic can be explained as accurately with the similar set of indicators mentioned in the literature. It also indicated that foot traffic can likewise be used as a measure of business performance because it shows similar dependencies with explanatory variables, including, for instance, net sales (Aboulola 2018).

Model performance	**Traditional**	**Novel**
Number of Observations:	121	121
Multiple R-Squared:	**0.295**	**0.296**
Adjusted R-Squared:	**0.231**	**0.211**
Akaike's Information Criterion (AICc):	303.636	311.087
Joint F-Statistic:	0.000	0.000

Table 4. Model performances of the models

Second, the Traditional model provided us a benchmark of assessing a set of novel indicators. Based on the model comparison (Table 4), it can be concluded that a model with novel indicators as LEED-certification and Tapestry Segment has performed very similar to Traditional model (adj. R^2 0.211, R^2 0.296 vs adj. R^2 0.231, R^2 0.295) and can be likewise applied for defining business performance of Starbucks stores.

Previous studies mentioned that site selection models, measure by net sales can be explained by about 29-31% using regression modelling (Aboulola 2018). Our novel model with a set of key variables explains about 30% of site performance, based on R^2 value.

Moreover, Tapestry data has only 67 categories which are presented in one single dataset. This implies a significantly fewer number of variables that do not require data transformation, in comparison to the 2289 unique sets provided by demographical data. Assessing from the model construction view, the novel model is more efficient for a decision-maker who deals with site selection predictions.

5.2. Hypothesis 1 – Sustainability and Business Performance

Business performance of selected sites is positively correlated with sustainability performance indicators. Higher sustainability score results in higher foot traffic.

The hypothesis has been tested by adding a novel sustainable indicator - LEED (sustainably certificated sites) in the site selection model (Appendix Model 2c). The result of the model has not shown any significance of LEED's relation to the business performance with p-value over 0.768 for LEED-certified stores and p-value 0.256 for LEED-certified stores with distinctions. The result also showed a positive slope, meaning that it is likely that a site would have a higher foot traffic, perform better if it was LEED-certified.

However, due to the low significance our hypothesis 1 that *business performance of selected sites is positively correlated with sustainability performance indicators* is FALSE. Higher sustainability scores had a positive correlation with business performance, but the results were not significant enough to have statistical sufficiency in evidence.

We can argue that probably more data is needed to calculate the dependency because our sample was presented with 21 LEED-certified buildings, meaning that only 18% of all Starbucks stores in Los Angeles had LEED-certification and less than 5% had certification with distinction. On the other hand, it is a real representation of the data and it might be so that our environmental indicator was not right for this analysis.

5.3. Hypothesis 2 – Socio-demographic segments and Business Performance

Tapestry segmentation is a significant indicator of business performance in site selection modeling. The presence of certain tapestry segments significantly correlates with the number of visitors per location.

Tapestry data was presented with 11 different lifestyles and 19 segments. The most common type in our sample was *"Uptown individuals"*, found in 58 sites out of 121, followed by "Next Wave". The most common segment was presented by *"Trendsetters"* which took place in ¼ of all sites. Nevertheless, the significant profiles were not only the ones that were among most representatives in the sample.

The model with Tapestry Life Modes (Appendix Model 2a) showed only one moderately significant Life Style variable and was considered as irrelevant for the study. On the other hand, the model with Tapestry Segments has shown a more detailed classification of socio-demographic profiles and provided higher significance between dependent and independent variables.

The Model with Tapestry segments (Appendix Model 2c) showed significant negative correlation with Tapestry segment Social Security Set (p-value 0.005) represented by the Life Mode of Senior Styles and Trendsetters (p-value 0.041) defined by Uptown individuals. The International Marketplace variable from Next Wave Life Mode had positive correlation (p-value 0.021) with business performance of Starbucks stores in Los Angeles.

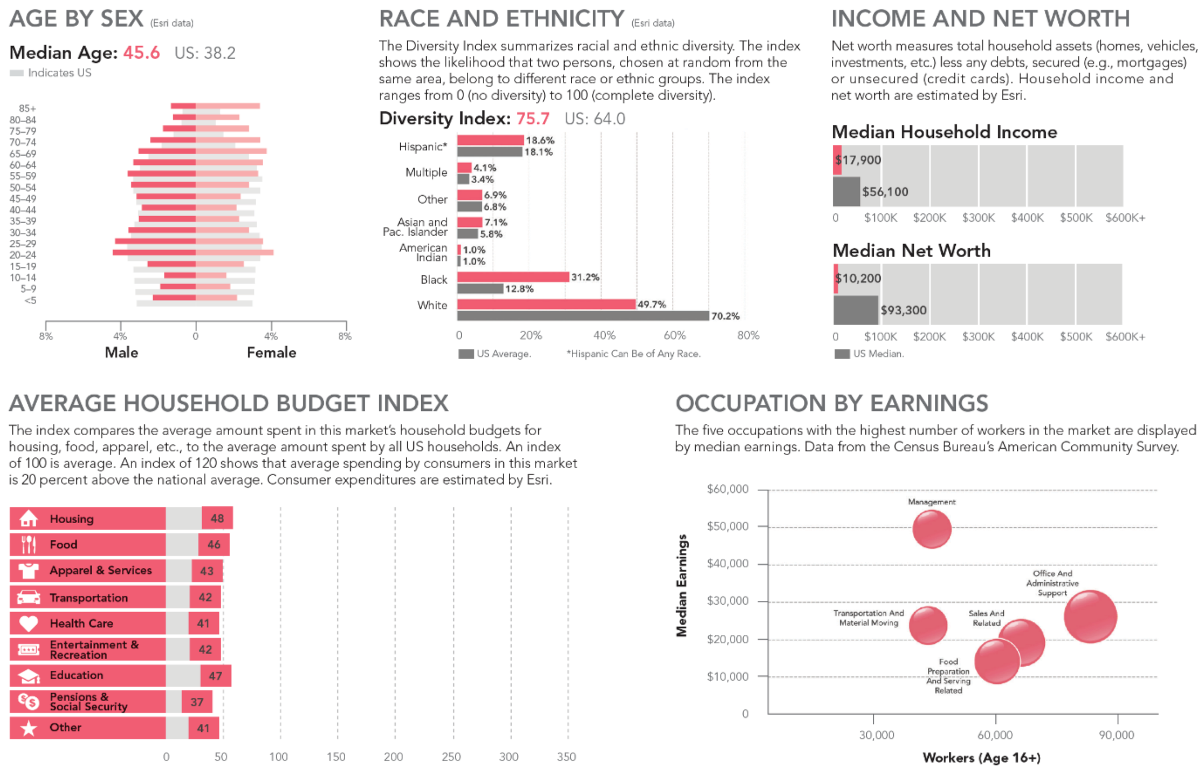


Figure 3. Social Security Set segment from https://doc.arcgis.com/en/esri-demographics/data/tapestry-segmentation.htm

By deeply analyzing the mentioned Tapestry profile, it can be concluded that Starbucks stores in Los Angeles should seriously consider how surroundings with the Social Security Set segment (Fugure 3), represented by the Senior Styles Life Mode, correlates with lower foot traffic activity. This indicates that these socioeconomic neighborhoods are not the most suitable for driving business performance. The segment is represented by the senior citizens living in metropolitan cities, with an average annual household income of 17,900 USD which is significantly below the US average of 56,100 USD. In addition, one-fourth of householders in this segment are aged 65 or older and are dependent on low, fixed incomes, primarily social security. Based on the profile description, it is not surprising that the relationship between business performance and presence of this segment is negative.

Another negative correlation was shown towards segment Trendsetters (Figure 4), represented by young individuals with high income and tendency to buy premium products. It might sound unexpected that young and active with high income (household income 63,100 USD) were not interested in Starbucks coffee. However, by looking closer at the data, this negative relation can be explained by the fact there might be other, more trendy coffee shops around Starbucks that these types of personas prefer. Market competition plays a significant role on business performance (p-value 0.034) and is positively correlated with the presence of Trendsetters (Appendix 7), meaning that there is a high demand for that socio-economic segment.

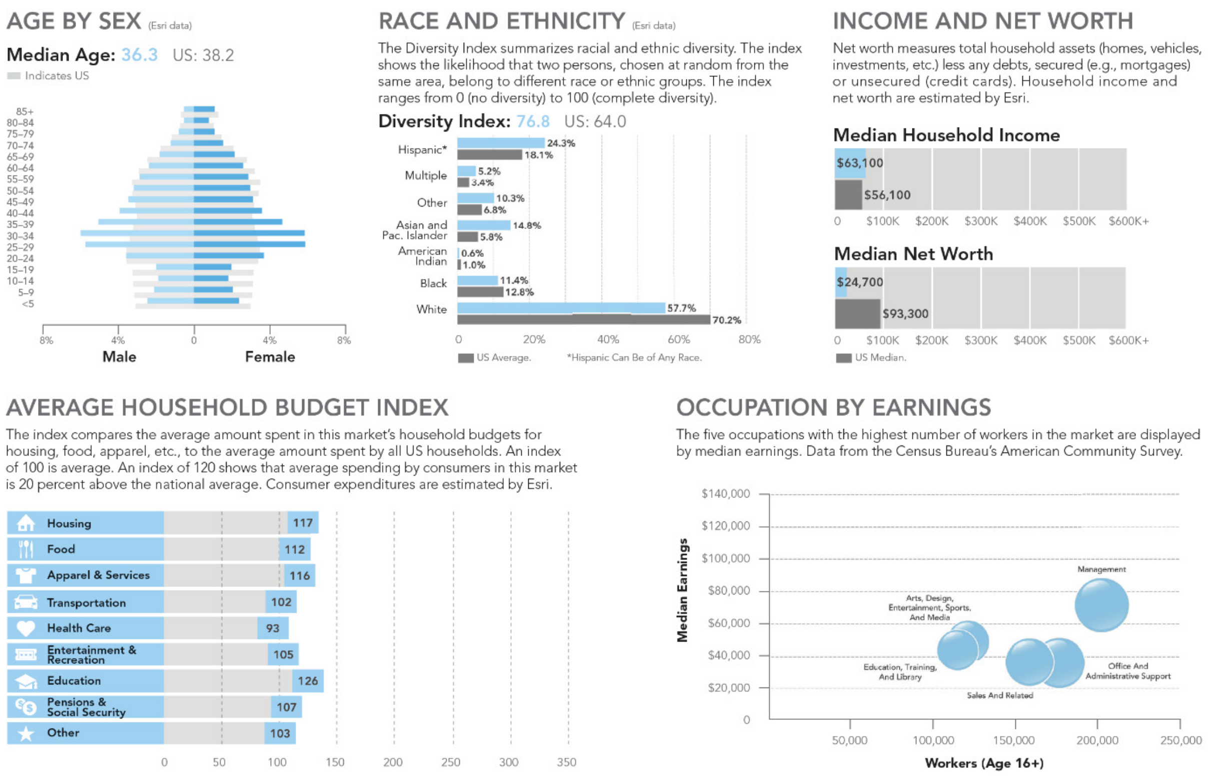


Figure 4. Trendsetters segment from https://doc.arcgis.com/en/esri-demographics/data/tapestry-segmentation.htm

The stores would likely show a higher business performance in the areas represented by the International Marketplace segment (Figure 5), from Next Way Life Mode. The segment defined by young diverse family market, 40% of which were born abroad with relatively low education. There are hardworking consumers with a slightly lower income that average. Preserving the environment and being in tune with nature are very important to them. As well as being connected in media using internet. One-fifth of workers commute using public transportation and walk or bike to work.

Population with lower education, higher diversity index and higher usage of public transportation were among the most significant indicators for successful Starbucks stores performance in the Traditional model (Appendix 3). The same indicators were represented by one single segment International Marketplace that also showed its positive correlation with business performance. This provides interesting insights that Tapestry segmentation profiles can be more efficient, in terms of the quantity of variables, than traditional demographical set of variables for site selection modeling.

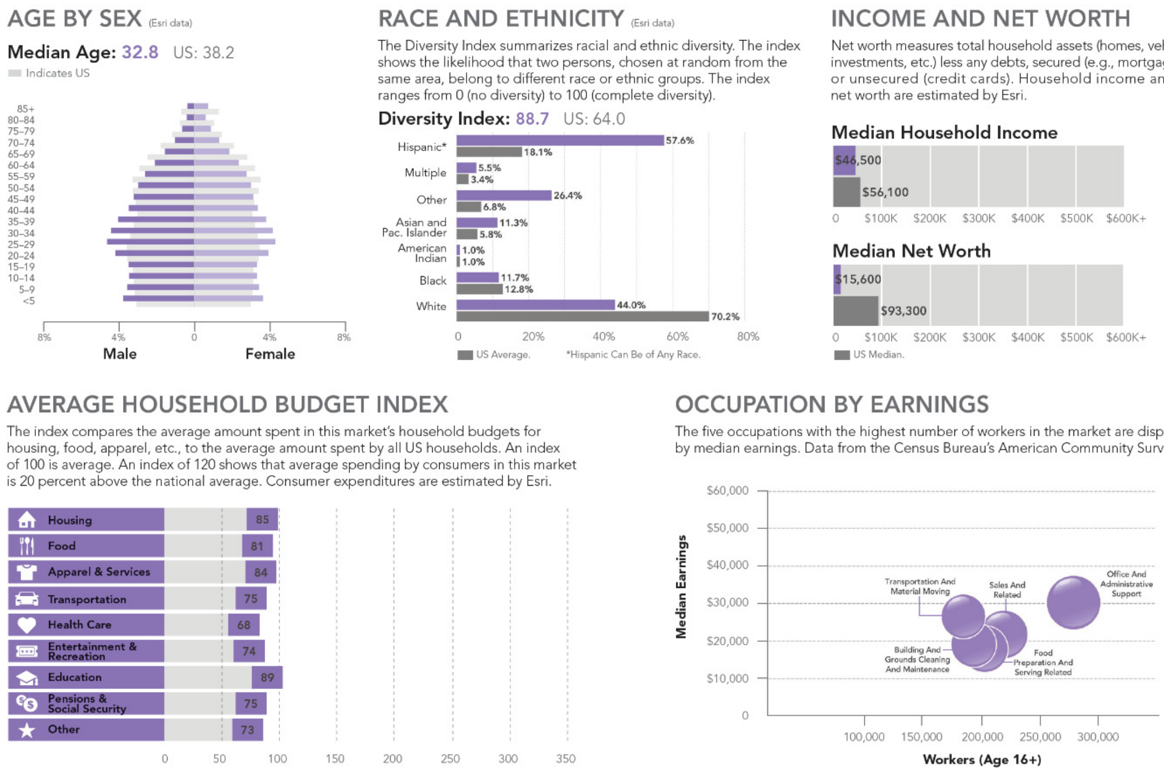


Figure 5. International Marketplace segment from https://doc.arcgis.com/en/esri-demographics/data/tapestry-segmentation.htm

Some Tapestry categories have shown a significant correlation with the business performance indicator, meaning that the higher or lower foot traffic can be partially explained by the presence or absence of certain Tapestry segments. Therefore, we can prove that our hypothesis *2 that business performance of selected sites could be explained by socio-demographic segments in site selection modeling and significantly correlates with number of customers* is TRUE.

5.4. Location driven decision-making process

By analyzing the overall site selection performance, we can clearly define the social, economic and demographic profiles of customers, as well as environmental surroundings of the store. The Huff's probabilistic model indicates that customers' choices were positively related to the size of shopping centers - this same indication is showed both of our models (p-value 0.001). That means that a bigger size of the location would most likely increase the concentration of foot traffic.

The proximity indicators as number of competitors around the site and number of stations, proved that location *matters*. Tobler statement that "everything is related to everything else, but near things are more related than distant things" is still relevant. According to Huff's theory the number of competitions is inverse to site performance. This statement has been validated with our results, as the number of competitor stores was one of the key indicators in the novel model (p-value 0.04). Crucially, it shows that less competition leads to better site's performance, as assumed.

Starting from 1931, the Law of Retail Gravitation by Reilly indicated that distance to road proximities is a key element in site performance. More recent studies (Chang 2018, Aboulola 2018) also brought up the importance of road proximity to a store's business performance. However, our study that focuses on environmental sustainability was intended to show the importance of sustainable transportation. We assumed that the distance to public stations would play a significant role as the distance to roads. Therefore, we used proximities to public stations instead of road distances to stores. The standard circle buffers were replaced with walking-distance polygons, called service areas that were calculated by the latest GIS-technique. The result proved our assumption. It has appeared that higher concentration of stations near the site indicates higher business performance. The number of stations was a major indicator of site performance (p-value of 0.003) for both traditional and novel models, which proves that the sustainable approach in site selection is relevant and should be considered in the future studies.

Consumption theory has stated that as household income increases, so does household spending and that people from better socio-economic backgrounds are more likely to actively shop and dine out, consequently driving local consumption up (Clower & Johnson 2017). Our study points out, however, that the single households' income measure does not always correlate positively with business performance of the retail store. In this view, one should look at more detailed socio-economical characteristics as well as customer preferences to predict the consummation of a retail segment.

6. Conclusion

Site selection remains a complex decision-making process, especially given constant changes in environment, increased new customers' needs and tastes, and new forms of competition. Developments in technology for data collection and analytics however provides more detailed information about customer preferences in a_geographical context. Crucially, this information can be beneficial for businesses and key industry-players in sharpening corporate decision-making and improving profitability. This study makes an empirical comparison between traditional site selection indicators (e.g. property size, distance to stations, market competition and demographic profiles) with novel site-selection indicators (e.g. environmental sustainability performance and socio-demographic Tapestry data) for Starbucks coffee shops. By testing a novel indicators and data collection and newest spatial analytical methods, we have been able to improve the traditional model, and lift the cons and pros from the alternative variables. Of importance, calculation of proximities (location-driven analysis) was tested with a modern walking-time proximity calculation technique rather than the traditional circle buffers. These results show a more accurate representation of the site's environment.

In doing so, we have proven how socio-demographic Tapestry data correlate with business performance indicators of small retail shops. The results showed that Tapestry segmentation stands as significant indicator of business performance in site selection modeling – specifically such segments of population as Social Security Set, Trendsetters and International Marketplace. Tapestry dataset clearly identifies economic, social and demographic features of customers as well as their preferences and is recommended to use in site selection models. Moreover, the novel site selection model become more user friendly and efficient considering the ease of application of 67 segments at most in comparison to thousands of demographic indicators. In addition, the study also demonstrates how higher quantity of public stations and lower competition leads to higher business performance.

The hypothesis that higher sustainability increased business performance has not been validated. Although, the results were non-significant, the trend between environmental performance and business performance for the site selection of coffee shops was positive. In addition, after analyzing the Tapestry customers profiles, there is an evidence that the typical persona who visits Starbucks cares about the environment. The non-significant results might be explained by the study limitations, one of which was a lack of LEED sustainability data.

In addition, this paper also suggests the significance of ethical considerations in site location. Although businesses often foreclose areas with large segments of low-income populations (often entailing old-age or racialized neighborhoods) as strategy of site selection optimization. These decision-making processes however perpetuate unequal gentrification without providing support for these neighborhoods. Instead, businesses ought to turn their attention to social sustainability, to ensure that business decisions do not entail community marginalization.

6.1. Limitations

The foot traffic data from mobile devices was used as a proxy to measure business performance of the sites. Due to the lack of access, we were not able to validate the actual counts of customers entered the stores.

The sustainability indicator represented by LEED-rankings has been found only in one fifth of locations. The small sample of sustainable dataset might have affected our results.

6.2. Suggestions for future studies

There are several findings that should be addressed to the scientific and business communities:

- Foot traffic is a satisfactory measure of the business performance of the small retail store, especially when there is a limited access to financial reports.
- Calculation of proximities by walking-time buffers provides more accurate representation of the site surroundings in comparison to traditional circle buffers.
- The store size is a significant indicator of sites performance, bigger stores generate higher demand.
- The walking proximities to public transportation and lower competition significantly correlates with higher business performance of small retail stores. Therefore, those indicators should be taken into consideration for future site selection models.
- Tapestry segmentation appears to be a significant and well-explanatory indicator of business performance in site selection modeling and its highly recommended to be applied in similar cases.

Future suggestions based on the study limitations:

1. It is worth considering financial information as a proxy for business performance.
2. Different metrics for foot traffic segmentation could be used. For instance, future research can look at hourly activities.
3. A longitudinal study to measure the effect of environmental indicators on site's performance, including testing for other proxies of sustainability.
4. Considering a different retail industry, broader sample of small retail stores, or other geographical areas.

As suggested in several studies by Aboula (2018), Amparo (2013), and Chen et al. (2018), it is important for companies to use novel site selection components as part of their decision-making models in order to stay competitive in the small retail business. Therefore, site selection parameters should be kept up to date with the most recent data collection techniques and the latest trends in GIS-technology. Moreover, the effect of sustainable business tends to show rather long-term effect, therefore, a longitudinal design study is worth considering in the future research.

This study enables the facilitation of new insights into alternative methods of location-allocation decision-making for small retail stores, as well as provides the valuable information for the future research that might be applicable on a broader scale or for other retail markets.

7. Bibliography

Aboulola, O 2018a, *A Literature Review of Spatial location Analysis for Retail Site Selection*, PhD Thesis, University of Jeddah, viewed 19 March 2020, <https://pdfs.semanticscholar.org/bfcc/c584dd903331ef6d4127f76b88204741283e.pdf>.

Aboulola, O 2018, *A New Model for Spatial Analysis Site Selection and Decision Making for Small Retail Facilities*, PhD Thesis, Claremont Graduate University, viewed 19 March 2020, <https://www.researchgate.net/publication/327384423_A_New_Model_for_Spatial_Analysis_Site_Selection_and_Decision_Making_for_Small_Retail_Facilities>.

Abrishami S, Kumar P & Nienaber W 2017, Smart Stores: A Scalable Foot Traffic Collection and Prediction System. In: Perner P. (eds) Advances in Data Mining. Applications and Theoretical Aspects. ICDM 2017. *Lecture Notes in Computer Science*, vol 10357. Springer, Cham. doi: https://doi.org/10.1007/978-3-319-62701-4_9.

Bergh, J, Ferrer-i-Carbonell A 1999. "*Economic theories of sustainable consumption,*" Serie Research Memoranda 0055, VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics.

Brown, S 2008, *Measures of Shape: Skewness and Kurtosis*, viewed 19 September 2017, <https://brownmath.com/stat/shape.htm>.

Brunauer, WA, Lang, S, Wechselberger, P & Bienert, S 2010, 'Additive Hedonic Regression Models with Spatial Scaling Factors: An Application for Rents in Vienna', *The Journal of Real Estate Finance and Economics,* vol. 41, no. 4, pp. 390-411.

Bryman, A 2006, '*Integrating Quantitative and Qualitative Research: How is It Done?*' Qualitative Research, vol. 6, no. 1, pp. 97-113.

Chang Xiaohui, Jiexun Li 2018, '*Business Performance Prediction in Location-based Social Commerce*', *Elsevier.* Expert Systems with Applications.

Chen L, Olhager J & Tang O 2012, 'Manufacturing Facility Location and Sustainability: A Literature Review and Research Agenda', *International Journal of Production Economics,* vol. 149, pp. 154-163.

Chen, J, Wang, J, Baležentis, T, Zagurskaitė, F, Streimikiene, D & Makutėnienė, D 2018, 'Multicriteria Approach Towards the Sustainable Selection of a Teahouse Location with Sensitivity Analysis', *Sustainability*, vol. 10, no. 8, pp. 2926-2942.

Choudhury, V & Karahanna, E 2008, 'The Relative Advantage of Electronic Channels: A Multidimensional View', MIS Quarterly, vol. 32, no. 1, pp. 179–200.

Clower, RW & Johnson, MB 2017, 'Income, Wealth, and The Theory of Consumption', in GR Ricci (ed), *Value, capital and growth, Routledge*, London, UK, pp. 45-96.

Cohen, J 1992, 'Statistical Power Analysis', *Current Directions in Psychological Science*, vol. 1, no. 88, pp. 98-101.

Denzer, AS & Hedges, KE 2011, 'The Limitations of LEED: A Case Study', *Journal of Green Building,* vol. 6, no. 1, pp. 25-33.

Dijkmans, C, Kerkhof, P & Beukeboom, CJ 2015, 'A Stage to Engage: Social Media Use and Corporate Reputation'*, Tourism Management*, vol. *47*, pp. 58-67.

Donner, H & Loh, T 2019, 'Does the Starbucks Effect Exist? Searching for A Relationship between Starbucks and Adjacent Rents', *Property Management,* viewed 15 December 2019, <https://doi-org.miman.bib.bth.se/10.1108/PM-01-2019-0004>.

ESRI 2020, *ESRI Demographics*, viewed 19 March 2020, <https://www.esri.com/en-us/arcgis/products/esri-demographics/overview>.

Farkas A 2009 '*Route/site Selection of Urban Transportation Facilities: An Integrated GIS/MCDM Approaches',* in *MEB 2009 – 7th International Conference on Management*, Enterprise and Benchmarking, Óbuda University, Budapest, pp. 5-6.

Fox, E, Postrel, S, & McLaughlin, A 2007, '*The Impact of Retail Location on Retailer Revenues: An Empirical Investigation'.*

Gallaugher, J & Ransbotham, S 2010, *'Social Media and Customer Dialog Management at Starbucks'*, *MIS Quarterly Executive*, vol. *9*, no. 4, pp. 197-212.

Glark, J.R., et al. 1997, 'A Model of Urban Forest Sustainability'. *Journal of Arboriculture*, 23, 17-30.

Grand View Research 2018, U.S. Hot Drinks Market Size, Share & Trends Analysis Report, by Product (Coffee, Tea, Others)*', Competitive Landscape, and Segment Forecasts, 2018 – 2025*, viewed 18 March 2020, <https://www.grandviewresearch.com/industry-analysis/us-hot-drinks-market>.

GreenBiz 2017, 'The Profitable Hospital System with Sustainability in its DNA', *GreenBiz,* viewed 17 March 2020, <https://www.greenbiz.com/article/profitable-hospital-system-sustainability-its-dna>.

Grekousis, G & Thomas, H 201, 'Comparison of Two Fuzzy Algorithms in Geodemographic Segmentation Analysis: The Fuzzy C-Means and Gustafson–Kessel Methods', Applied Geography, vol. *34*, pp. 125-136.

GRR 2020, *Global risks report 2020: executive summary*, viewed 17 March 2020, <http://reports.weforum.org/global-risks-report-2020/executive-summary/>.

Han, H, Nguyen, HN, Song, H, Chua, BL, Lee, S & Kim, W 2018, 'Drivers of Brand Loyalty in the Chain Coffee Shop Industry', *International Journal of Hospitality Management*, vol. 72, pp. 86-97.

Henver, A, & Chatterjee, S 2010b 'Design Science Research in Information Systems', in *Design Research in Information Systems*, vol. 22, Springer, Boston, MA, pp. 9-22.

Infogroup, 2020, ESRI's business data provider viewed 21 March 2020, <https://www.infogroup.com//>.

Huff, D, L 1963 'A probabilistic analysis of shopping center trade areas', *Land Economics*, vol. 39, no. 1, pp. 81-90.

Jang, Y, J, Kim, W, G, & Lee, H, Y 2015, 'Coffee shop consumers' emotional attachment and loyalty to green stores: The moderating role of green consciousness', *International Journal of Hospitality Management*, vol. 44, pp. 146-156. Retrieved from http://www.sciencedirect.com/science/article/pii/S0278431914001509. doi:https://doi.org/10.1016/j.ijhm.2014.10.001

Kang, HO, Lee, SM, & Kim, YH 2010, 'Menu analysis for coffee shop operation: using activity-based costing', *International CHRIE Conference-Refereed Track*, vol. 2010, pp. 1-8.

Kelly, J, P, Freeman, D, C, & Emlen, J, M 1993, 'Competitive impact model for site selection: the impact of competition, sales generators and own store cannibalization', *The International Review of Retail, Distribution and Consumer Research*, vol. 3, no. 3, pp. 237-259. Retrieved from https://doi.org/10.1080/09593969300000017. doi:10.1080/09593969300000017

Larson DG 2019, *Leed for retail, growing a sustainable brand*, viewed 27 February 2020, <https://www.larsondesigngroup.com/leed-for-retail/>.

Lees, L, Slater, T & Wyly, E 2010, *The gentrification reader*, Routledge, London.

Li, Y, & Liu, L 2012, 'Assessing the impact of retail location on store performance: A comparison of Wal-Mart and Kmart stores in Cincinnati', *Applied Geography*, vol. 32, no. 2, pp. 591-600. Retrieved from http://www.sciencedirect.com/science/article/pii/S014362281100141X. doi:https://doi.org/10.1016/j.apgeog.2011.07.006

Lord JD, Wright DB. Competition and location strategy in branch banking: spatial avoidance or clustering. Urban Geograhy 1981;2:3:189–200.

Mankiw, NG 2011, *Principles of economics*, 6 edn, Cengage Learning, Boston, MA.

Nicasio, F 2018, *Retail metrics: 14 essential KPIs for tracking your business' performance*, viewed 21 December 2019, <https://www.vendhq.com/blog/retail-metrics-and-kpis/>.

Obaidullah Jan, 2019, *Revenue per employee*, viewed 21 March 2020, <https://xplaind.com/819132/revenue-per-employee/>.

Queiros A, Gonzalez GH 2019, *Railway mobility hubs: a feature-based investment return analysis,* MBA Thesis, Blekinge Institute of Technology, viewed 5 March, 2020, < http://www.diva-portal.org/smash/get/diva2:1331462/FULLTEXT02.pdf>.

Reinartz, W.J. & Kumar, V. (1999). Store-, market-, and consumer-charactersitics: the drivers of store performance. Marketing Letter, 10(1), 5-22. (edit)

Rezaee, Z 2016, 'Business sustainability research: a theoretical and integrated perspective', *Journal of Accounting Literature*, vol. *36*, pp. 48-64.

Reilly, W J 1931. The law of retail gravitation: WJ Reilly.

Rendtorff, J 2019, 'Corporate social responsibility, sustainability, and stakeholder management', in *Philosophy of management and sustainability: rethinking business ethics and social responsibility in sustainable development*, Emerald Publishing Limited, Bingley, UK, pp. 43-52.

Roig-Tierno, N, Baviera-Puig, A, Buitrago-Vera, J & Mas-Verdu, F 2013, 'The retail site location decision process using GIS and the analytical hierarchy process', *Applied Geography*, vol. 40, pp. 191-198.

San Martin, D, Orive, M, Martínez, E, Iñarra, B, Ramos, S, González, N, de Salas, AG, Vázquez, L & Zufía, J 2017, 'Decision making supporting tool combining AHP method with GIS for implementing food waste valorisation strategies', Waste and Biomass Valorization, vol. 8, no. 5, pp. 1555-1567.

Satani, N, Uchida, A, Deguchi, A, Ohgai, A, Sato, S, & Hagishima, S 1998, 'Commercial facility location model using multiple regression analysis', *Computers, Environment and Urban Systems*, vol. 22, no. 3, pp. 219-240.

Saunders, MN, Lewis, P & Thornhill, A 2016, *Research methods for business students*,7th edn, Pearson, London, UK.

Shelton, B 2016, 'The retail site selection process', *Buxton*, viewed 21 March 2020, <https://www.buxtonco.com/pages/the-retail-site-selection-process/>.

Skewness graph 2018. The California State University Social Science Research and Instructional Center, viewed 21 March 2020, <https://ssric.org/node/495/>.

Schmidt CG, Lee Y. An intermetropolitan analysis of retail locational patterns and competitive processes. Social Science Journal1979;16:43–53

Son, K, Choi, K, Woods, P and Park, YJ 2012 'Urban sustainability predictive model using GIS: appraised land value versus LEED sustainable site credits', *Journal of Construction Engineering and Management*, vol. 138, no. 9, pp.1107-1112.

Starbucks 2018, Global social impact report 2018, viewed 27 February 2020, https://stories.starbucks.com/uploads/2019/08/2018-Global-Social-Impact-Report.pdf

Statista 2019, *Hot drinks report 2019*, viewed 16 December 2019, <https://www.statista.com/outlook/30000000/109/hot-drinks/united-states>.

Steingold, FS 2011, *Legal guide for starting & running a small business*, 12th edn, NOLO, Berkeley, CA.

Sustainability Reports 2019, *86% of S&P 500 Index® Companies Publish Sustainability / Responsibility Reports in 2018*, viewed 4 March 2020, <https://www.sustainability-reports.com/86-of-sp-500-index-companies-publish-sustainability-responsibility-reports-in-2018/>.

Taryn R & Zhou Y 2019, *Environmental Performance, Customer Satisfaction and Profitability: A study among Large U.S. Companies*. MDPI. Sustainability.

Terouhid, S A, Ries, R, & Fard, M 2012, 'Towards Sustainable Facility Location – A Literature Review', *Journal of Sustainable Development*, vol. 5. doi:10.5539/jsd.v5n7p18

Tkaczyk, C 2016, 'The 9 key ingredients for business success', *Forbes*, 19 February, viewed 5 July 2017, <http://fortune.com/2016/02/19/9-key-ingredients-to-business-success/>.

Tobler, WR 1970, 'A computer movie simulating urban growth in the Detroit region', *Economic Geography*, vol. 46, p. 234 .

Tong, C H, & Tong, L I 2006, 'Exploring the cornerstones of Wal-mart's success and competitiveness', *Competitiveness Review: An International Business Journal*.

Towncharts 2020, *Retail site selection*, viewed 21 December 2019, <https://www.towncharts.com/Retail-Site-Selection.html>.

Turhan, G 2013, 'Literature review on selection criteria of store location based on performance measures', *Procedia - Social and Behavioral Sciences*, vol. 99, pp. 391-402.

Tzeng, G-H, Teng, M-H, Chen, J-J, Opricovic, S 2002, 'Multicriteria selection for a restaurant location in Taipei', *International Journal of Hospitality Management*, vol. 21, no. 2, pp. 171-187.

Ubermedia 2020, Uber Media, Inc., Pasadena, CA, viewed 15 March 2020, <https://ubermedia.com/>.

USGBC 2019, *LEED In Motion Retail 2019*, viewed 3 March 2020, <https://www.usgbc.org/resources>.

USGBC Projects 2019, *U.S. Green Building Council projects database*, viewed 3 March 2020, <https://www.usgbc.org/projects>.

Vlachou, C & Iakovidou, O 2016 'The evolution of studies on business location factors', *Journal of Developmental Entrepreneurship*, vol. 20, no. 4, 1550023.

Wang, F, Chen, C, Xiu, C & Zhang, P 2014, 'Location analysis of retail stores in Changchun, China: A street centrality perspective', *Cities*, vol. 41, pp. 54-63. Retrieved from http://www.sciencedirect.com/science/article/pii/S0264275114000663. doi:https://doi.org/10.1016/j.cities.2014.05.005

Wang, L, Fan, H & Wang, Y, 2018, 'Site selection of retail shops based on spatial accessibility and hybrid BP neural network', *ISPRS International Journal of Geo-Information*, vol. 7, no. 6, pp. 202-222.

Zentes J, Morschett D & Schramm-Klein H, 2011, *Store Location – Trading Area Analysis and Site Selection*. In: Strategic Retail Management. Gabler Verlag, doi: https://doi.org/10.1007/978-3-8349-6740-4_11

7. Appendix

I. Dataset with default variables used for the analysis

Sample of the dataset used for the study. The full dataset is available upon the request.

SiteID	FootTraffic	StoreSize	TotalPopulation	IncomePerCapita	MedianAge	BachelorDegree	TotalHouseholds	Coffee	TotalBusinesses	Competition	Stations	TapestryID	LEED
1	6986	2	9114	11517	30.4	612	2944	150225	272	0	20	62	0
2	4374	2	1572	18872	33.3	87	477	40407	79	0	10	51	0
3	2211	2	1139	55795	45	301	431	84994	365	0	3	7	0
4	2922	2	1856	53651	39.1	495	892	143729	114	0	2	12	0
5	6247	2	892	33764	42.4	204	350	41840	60	0	9	33	0
6	4158	1	2042	44611	36.2	587	979	125687	200	0	13	12	0
7	5514	2	2884	21555	36.6	724	1039	80734	147	0	10	60	0
8	6236	2	1616	11346	27.3	89	368	26950	63	0	10	61	0
9	10251	2	715	9694	36.1	51	299	37255	165	0	18	11	1
10	2968	2	94	15545	39.8	11	44	2209	235	0	34	45	0
11	1687	2	1642	55840	33.9	325	628	128860	296	0	12	10	0
12	423	1	1460	42494	37.6	335	689	94084	159	0	11	12	0
13	4024	2	3459	28889	24.2	465	2006	145225	642	0	8	67	0
14	3032	2	2630	28048	34.5	357	839	98654	112	0	10	60	0
15	4964	2	1492	12228	28.7	40	385	27331	29	0	4	61	0
16	1616	2	427	93537	44.7	159	216	52047	109	0	7	6	0
17	3037	2	5454	30809	32.7	1144	3144	291889	1198	0	22	12	0
18	3522	2	1936	25848	38.5	453	710	72521	73	0	9	33	0
19	9721	2	2135	26930	35.7	329	731	83777	105	0	12	60	0
20	9496	1	685	37675	45.9	133	337	35208	51	0	9	33	0
21	8135	2	930	22239	38.4	82	308	30213	43	0	12	56	0
22	10593	2	3952	76063	34.2	1339	2431	416192	647	0	21	11	0
23	1339	2	3424	12176	47.7	420	1262	64481	354	0	24	45	1
24	5681	2	1531	10407	26.7	32	347	23569	97	0	7	61	1
25	19683	1	1331	10670	29.4	57	331	20266	117	0	16	61	1

2. Model 1a – All Traditional variables

Explanatory Variable(s) Income_pc_log;Tot_Households_log;Tot_Business_log;Ed_Bachelor_Log;Tot_Crime_log;Spendings_Coffee_Log;Median_Age_Log;Diversity_Index;Unemployment_Rate;Competition_log;Stations_log;Store_size_2

```
Running script GeneralizedLinearRegression...
----------------------- Summary of GLR Results [Model Type: Continuous (Gaussian/OLS)] -----------------------
--------------------------------------------------------------------------------------------------------------
            Variable Coefficient [a] StdError t-Statistic Probability [b] Robust_SE  Robust_t Robust_Pr [b]    VIF [c]
           Intercept       11.853122 2.490376    4.759570       0.000007*  2.523670  4.696779     0.000009*   --------
       INCOME_PC_LOG       -0.321124 0.351323   -0.914040       0.362724   0.342142 -0.938567     0.350039   10.419453
  TOT_HOUSEHOLDS_LOG       -0.032616 0.543896   -0.059967       0.952286   0.494560 -0.065949     0.947533  106.174710
    TOT_BUSINESS_LOG       -0.213260 0.106851   -1.995868       0.048463*  0.096574 -2.208244     0.029334*   3.754233
     ED_BACHELOR_LOG        0.193781 0.198531    0.976070       0.331204   0.146359  1.324006     0.188301   16.119031
       TOT_CRIME_LOG        0.438550 0.238054    1.842226       0.068188   0.232268  1.888121     0.061695    1.991121
SPENDINGS_COFFEE_LOG       -0.624709 0.452666   -1.380065       0.170423   0.410116 -1.523251     0.130627   76.795908
      MEDIAN_AGE_LOG        2.592493 1.247059    2.078886       0.039992*  1.237255  2.095359     0.038472*  25.053763
     DIVERSITY_INDEX       -0.013555 0.006807   -1.991285       0.048973*  0.007091 -1.911591     0.058579    2.137585
   UNEMPLOYMENT_RATE       -0.051282 0.026992   -1.899892       0.060115   0.021675 -2.365980     0.019758*   1.853426
     COMPETITION_LOG       -0.095419 0.152248   -0.626732       0.532158   0.128608 -0.741935     0.459732    1.758888
        STATIONS_LOG        0.299573 0.118743    2.522879       0.013091*  0.122441  2.446667     0.016025*   2.364004
        STORE_SIZE_2        0.619847 0.196694    3.151325       0.002109*  0.202998  3.053470     0.002853*   1.117766
--------------------------------------------------------------------------------------------------------------
----------------------------------------------- GLR Diagnostics -----------------------------------------------
---------------------------------------------------------------------------------------------------------------
Input Features:             GRL_analysis_200423   Dependent Variable:                                FOOT_TRAFFIC_LOG
Number of Observations:                     121   Akaike's Information Criterion (AICc) [d]:               304.338852
Multiple R-Squared [d]:                0.319697   Adjusted R-Squared [d]:                                    0.244108
Joint F-Statistic [e]:                 4.229407   Prob(>F), (12,108) degrees of freedom:                     0.000019*
Joint Wald Statistic [e]:             61.352151   Prob(>chi-squared), (12) degrees of freedom:               0.000000*
Koenker (BP) Statistic [f]:           10.539881   Prob(>chi-squared), (12) degrees of freedom:               0.568707
Jarque-Bera Statistic [g]:             1.944108   Prob(>chi-squared), (2) degrees of freedom:                0.378305
---------------------------------------------------------------------------------------------------------------
Notes on Interpretation
 *  An asterisk next to a number indicates a statistically significant p-value (p < 0.01).
```

3. Model 1b – after deleting variables with high VIF*

Explanatory Variable(s) Income_pc_log;Tot_Business_log;Ed_Bachelor_Log;Tot_Crime_log;Median_Age_Log;Diversity_Index;Unemployment_Rate;Competition_log;Stations_log;Store_size_2

VIF values above 7.5 should not cause any jeopardizing of the model. Total Households and Spending on Coffee were gradually removed from the model.

```
*Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among
explanatory variables.

Running script GeneralizedLinearRegression...
--------------------- Summary of GLR Results [Model Type: Continuous (Gaussian/OLS)] ---------------------
---------------------------------------------------------------------------------------------------------
         Variable Coefficient [a] StdError t-Statistic Probability [b] Robust_SE  Robust_t Robust_Pr [b]  VIF [c]
        Intercept       12.613150 2.450157    5.147894       0.000002*  2.617205  4.819320     0.000006* --------
    INCOME_PC_LOG       -0.611102 0.207622   -2.943333       0.003964*  0.228206 -2.677845     0.008542* 3.574871
 TOT_BUSINESS_LOG       -0.209460 0.096173   -2.177954       0.031538*  0.086063 -2.433804     0.016544* 2.987802
  ED_BACHELOR_LOG       -0.027223 0.134267   -0.202756       0.839700   0.099449 -0.273742     0.784801  7.242716
    TOT_CRIME_LOG        0.360310 0.236615    1.522768       0.130695   0.238580  1.510224     0.133862  1.932462
   MEDIAN_AGE_LOG        0.762310 0.613900    1.241750       0.216972   0.498952  1.527822     0.129436  5.964517
  DIVERSITY_INDEX       -0.012846 0.006859   -1.873021       0.063722   0.007215 -1.780498     0.077760  2.131736
UNEMPLOYMENT_RATE       -0.044569 0.026606   -1.675174       0.096748   0.022514 -1.979647     0.050239  1.768986
  COMPETITION_LOG       -0.120921 0.151277   -0.799335       0.425812   0.127558 -0.947968     0.345216  1.705931
     STATIONS_LOG        0.277787 0.115024    2.415030       0.017375*  0.115712  2.400678     0.018035* 2.179194
     STORE_SIZE_2        0.597604 0.197603    3.024273       0.003106*  0.209424  2.853557     0.005167* 1.108243
---------------------------------------------------------------------------------------------------------
----------------------------------------------- GLR Diagnostics -----------------------------------------------
---------------------------------------------------------------------------------------------------------------
Input Features:             GRL_analysis_200423   Dependent Variable:                                FOOT_TRAFFIC_LOG
Number of Observations:                     121   Akaike's Information Criterion (AICc) [d]:               303.636212
Multiple R-Squared [d]:                0.294674   Adjusted R-Squared [d]:                                    0.230554
Joint F-Statistic [e]:                 4.595635   Prob(>F), (10,110) degrees of freedom:                     0.000019*
Joint Wald Statistic [e]:             49.070914   Prob(>chi-squared), (10) degrees of freedom:               0.000000*
Koenker (BP) Statistic [f]:           11.295700   Prob(>chi-squared), (10) degrees of freedom:               0.334949
Jarque-Bera Statistic [g]:             2.991196   Prob(>chi-squared), (2) degrees of freedom:                0.224115
---------------------------------------------------------------------------------------------------------------
Notes on Interpretation
 *  An asterisk next to a number indicates a statistically significant p-value (p < 0.01).
```

4. Model 2a Tapestry Life Modes and LEED – corrected

Explanatory Variable(s) Competition_log;Stations_log;Store_size_2;LEED_1;LEED_2;T_Upscale_Avenues;T_Uptown_Individuals;T_Scholars_Patriots;T_Middle_Ground;T_Next_Wave;T_Senior_Styles

```
---------------------- Summary of GLR Results [Model Type: Continuous (Gaussian/OLS)] ----------------------
------------------------------------------------------------------------------------------------------------
           Variable Coefficient [a] StdError t-Statistic Probability [b] Robust_SE  Robust_t Robust_Pr [b]  VIF [c]
          Intercept        6.922290 0.357406   19.368116       0.000000*  0.337435 20.514435     0.000000* --------
    COMPETITION_LOG       -0.303016 0.133294   -2.273289       0.024961*  0.126677 -2.392037     0.018459* 1.279729
       STATIONS_LOG        0.288125 0.098446    2.926721       0.004172*  0.098083  2.937548     0.004040* 1.542401
       STORE_SIZE_2        0.666832 0.202751    3.288921       0.001362*  0.223803  2.979552     0.003563* 1.127347
             LEED_1        0.044098 0.224179    0.196710       0.844420   0.248284  0.177612     0.859354  1.118786
             LEED_2        0.448109 0.423292    1.058629       0.292104   0.415031  1.079700     0.282654  1.308522
  T_UPSCALE_AVENUES       -0.176543 0.363716   -0.485386       0.628383   0.350004 -0.504402     0.615003  2.015503
T_UPTOWN_INDIVIDUALS      -0.208008 0.278905   -0.745802       0.457387   0.253236 -0.821397     0.413206  3.578934
 T_SCHOLARS_PATRIOTS      -0.018698 0.448849   -0.041657       0.966842   0.337015 -0.055480     0.955851  1.750345
    T_MIDDLE_GROUND       -0.026819 0.393163   -0.068214       0.945733   0.427254 -0.062771     0.950057  1.759494
        T_NEXT_WAVE        0.516326 0.322499    1.601016       0.112276   0.284070  1.817603     0.071875  2.645337
    T_SENIOR_STYLES       -0.745647 0.425517   -1.752330       0.082535   0.381234 -1.955877     0.053038  2.060993
------------------------------------------------------------------------------------------------------------
----------------------------------------------- GLR Diagnostics -----------------------------------------------
---------------------------------------------------------------------------------------------------------------
Input Features:             GRL_analysis_200423   Dependent Variable:                               FOOT_TRAFFIC_LOG
Number of Observations:                     121   Akaike's Information Criterion (AICc) [d]:              309.200672
Multiple R-Squared [d]:                0.276661   Adjusted R-Squared [d]:                                   0.203663
Joint F-Statistic [e]:                 3.790001   Prob(>F), (11,109) degrees of freedom:                    0.000079*
Joint Wald Statistic [e]:             62.078426   Prob(>chi-squared), (11) degrees of freedom:              0.000000*
Koenker (BP) Statistic [f]:           22.284970   Prob(>chi-squared), (11) degrees of freedom:              0.022256*
Jarque-Bera Statistic [g]:             3.733967   Prob(>chi-squared), (2) degrees of freedom:               0.154589
---------------------------------------------------------------------------------------------------------------
Notes on Interpretation
 *  An asterisk next to a number indicates a statistically significant p-value (p < 0.01).
```

Life Modes with less than 5% presence in the sample were excluded from the model due to a small number of observations in these categories (Tapestry table).

5. Model 2b All Tapesty Segments

Summary of GLR Results [Model Type: Continuous (Gaussian/OLS)]

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	6.812251	0.347622	19.596748	0.000000*	0.293765	23.189464	0.000000*	--------
COMPETITION_LOG	-0.290415	0.135897	-2.137017	0.034913*	0.123974	-2.342554	0.021030*	1.322728
STATIONS_LOG	0.340450	0.109981	3.095545	0.002524*	0.100036	3.403283	0.000951*	1.914186
STORE_SIZE_2	0.668019	0.211127	3.164068	0.002041*	0.220819	3.025195	0.003128*	1.215544
T6	0.016997	0.430925	0.039443	0.968607	0.404297	0.042041	0.966540	1.604275
(T7)	-0.321484	0.468742	-0.685844	0.494320	0.511225	-0.628851	0.530813	1.287477
(T9)	-0.236869	0.862724	-0.274560	0.784198	0.260514	-0.909237	0.365299	1.118280
T10	0.065820	0.340211	0.193468	0.846965	0.277250	0.237402	0.812809	1.753509
T11	-0.087541	0.299468	-0.292322	0.770622	0.248407	-0.352411	0.725244	2.081785
T12	-0.388354	0.274907	-1.412676	0.160715	0.247648	-1.568165	0.119860	2.524538
T33	-0.138367	0.392963	-0.352112	0.725467	0.429049	-0.322498	0.747724	1.542880
T45	-0.699648	0.392688	-1.781689	0.077695	0.330535	-2.116710	0.036642*	1.745377
T60	0.581675	0.351643	1.654164	0.101088	0.275637	2.110291	0.037204*	1.718510
T61	0.496654	0.408533	1.215701	0.226826	0.403073	1.232169	0.220639	1.441882
(T62)	0.293331	0.478071	0.613573	0.540826	0.395962	0.740806	0.460460	1.339234
(T67)	0.395937	0.468637	0.844869	0.400097	0.404083	0.979840	0.329410	1.286900

GLR Diagnostics

Input Features:	GRL_analysis_200423	Dependent Variable:	FOOT_TRAFFIC_LOG
Number of Observations:	121	Akaike's Information Criterion (AICc) [d]:	315.898497
Multiple R-Squared [d]:	0.299268	Adjusted R-Squared [d]:	0.199163
Joint F-Statistic [e]:	2.989551	Prob(>F), (15,105) degrees of freedom:	0.000376*
Joint Wald Statistic [e]:	249.481848	Prob(>chi-squared), (15) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	24.466541	Prob(>chi-squared), (15) degrees of freedom:	0.057583
Jarque-Bera Statistic [g]:	1.760674	Prob(>chi-squared), (2) degrees of freedom:	0.414643

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

Variables T7, T9, T62, T67 (marked in brackets) were excluded from the final model due to small amount of counts in the sample.

6. Model 2c Tapestry Segments with LEED

Explanatory Variable(s) Competition_log;Stations_log;Store_size_2;LEED_1;LEED_2;T6;T10;T11;T12;T33;T45;T60;T61

Summary of GLR Results [Model Type: Continuous (Gaussian/OLS)]

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	6.751998	0.307705	21.943077	0.000000*	0.282214	23.925138	0.000000*	--------
COMPETITION_LOG	-0.273547	0.134231	-2.037881	0.044026*	0.124738	-2.192967	0.030468*	1.309007
STATIONS_LOG	0.366499	0.107026	3.424394	0.000882*	0.095909	3.821316	0.000229*	1.838725
STORE_SIZE_2	0.664795	0.205579	3.233777	0.001632*	0.217143	3.061554	0.002789*	1.169032
LEED_1	0.069307	0.223391	0.310251	0.756980	0.229120	0.302494	0.762869	1.120540
LEED_2	0.483136	0.393803	1.226848	0.222574	0.423593	1.140566	0.256595	1.142347
T6	0.062299	0.406195	0.153372	0.878390	0.388223	0.160472	0.872808	1.445878
T10	0.081161	0.305062	0.266050	0.790717	0.250536	0.323952	0.746614	1.430123
T11	-0.104938	0.250062	-0.419649	0.675592	0.211863	-0.495311	0.621402	1.472368
T12	-0.425466	0.220606	-1.928623	0.056427	0.205371	-2.071696	0.040694*	1.649046
T33	-0.124397	0.356277	-0.349159	0.727664	0.420018	-0.296171	0.767679	1.286446
T45	-0.888228	0.360645	-2.462886	0.015370*	0.309246	-2.872236	0.004916*	1.493281
T60	0.569521	0.307272	1.853474	0.066572	0.242935	2.344336	0.020899*	1.331012
T61	0.478779	0.373271	1.282658	0.202387	0.379131	1.262834	0.209396	1.220986

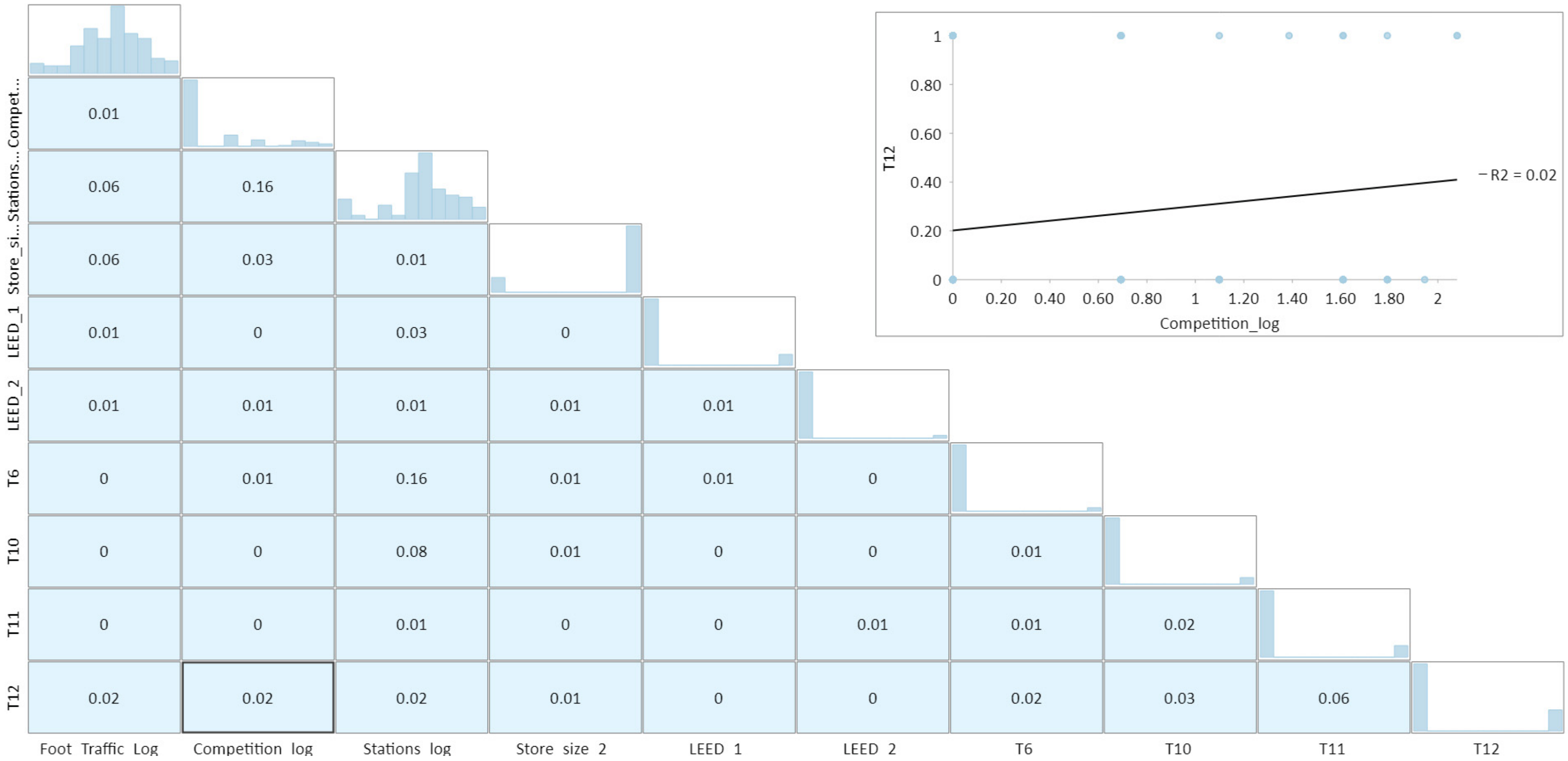
GLR Diagnostics

Input Features:	GRL_analysis_200423	Dependent Variable:	FOOT_TRAFFIC_LOG
Number of Observations:	121	Akaike's Information Criterion (AICc) [d]:	311.087534
Multiple R-Squared [d]:	0.296021	Adjusted R-Squared [d]:	0.210491
Joint F-Statistic [e]:	3.461012	Prob(>F), (13,107) degrees of freedom:	0.000113*
Joint Wald Statistic [e]:	72.051379	Prob(>chi-squared), (13) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	21.304283	Prob(>chi-squared), (13) degrees of freedom:	0.067124
Jarque-Bera Statistic [g]:	3.951448	Prob(>chi-squared), (2) degrees of freedom:	0.138661

Notes on Interpretation
* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

T6 - Urban Chic; T10 - Laptops and Lattes; T11 - Metro Renters; T12 – Trendsetters; T33 - City Lights; T45 - Social Security Set; T60 - International Marketplace; T61 - Las Casas

7. Correlation Matrix - relationships between variables (Novel Model 2c)



8. Tapestry Segmentation and LifeModes

LifeMode 1 Affluent Estates

- *Established wealth—educated, well-traveled married couples*
- *Accustomed to "more": less than 10% of all households, with 20% of household income*
- *Homeowners (almost 90%), with mortgages (65.2%)*
- *Married couple families with children ranging from grade school to college*
- *Expect quality; invest in time-saving services*
- *Participate actively in their communities*
- *Active in sports and enthusiastic travelers*

Segmetns: 1A Top Tier; 1B Professional Pride; 1C Boomburbs; 1D Savvy Suburbanites; 1E Exurbanites

LifeMode 2 Upscale Avenues

- Prosperous married couples living in older suburban enclaves
- Ambitious and hard-working
- Homeowners (70%) prefer denser, more urban settings with older homes and a large share of townhomes
- A more diverse population, primarily married couples, many with older children
- Financially responsible, but still indulge in casino gambling and lotto tickets
- Serious shoppers, from Nordstrom's to Marshalls or DSW, that appreciate quality, and bargains
- Active in fitness pursuits like bicycling, jogging, yoga, and hiking
- Subscribe to premium movie channels like HBO and Starz

Segmetns: *2A Urban Chic; 2B Pleasantville; 2C Pacific Heights; 2D Enterprising Professionals*

LifeMode 3 Uptown Individuals

- Young, successful singles in the city
- Intelligent (best educated market), hard-working (highest rate of labor force participation) and averse to traditional commitments of marriage and home ownership
- Urban denizens, partial to city life, high-rise apartments and uptown neighborhoods
- Prefer credit cards over debit cards, while paying down student loans
- Green and generous to environmental, cultural and political organizations
- Internet dependent, from social connections to shopping for fashion, tracking investments, making travel arrangements, and watching television and movies
- Adventurous and open to new experiences and places

Segmetns: *3A Laptops and Lattes; 3B Metro Renters; 3C Trendsetters*

LifeMode 4 Family Landscapes

- Successful young families in their first homes
- Non-diverse, prosperous married-couple families, residing in suburban or semirural areas with a low vacancy rate (second lowest)
- Homeowners (79%) with mortgages (second highest %), living in newer single-family homes, with median home value slightly higher than the U.S.
- Two workers in the family, contributing to the second highest labor force participation rate, as well as low unemployment
- Do-it-yourselfers, who work on home improvement projects, as well as their lawns and gardens
- Sports enthusiasts, typically owning newer sedans or SUVs, dogs, and savings accounts/plans, comfortable with the latest technology
- Eat out frequently at fast food or family restaurants to accommodate their busy lifestyle
- Especially enjoy bowling, swimming, playing golf, playing video games, watching movies rented via Redbox, and taking trips to a zoo or theme park

Segmetns: *4A Soccer Moms; 4B Home Improvement; 4C Middleburg*

LifeMode 5 GenXurban

- Gen X in middle age; families with fewer kids and a mortgage
- Second largest Tapestry group, comprised of Gen X married couples, and a growing population of retirees
- About a fifth of residents are 65 or older; about a fourth of households have retirement income
- Own older single-family homes in urban areas, with 1 or 2 vehicles
- Live and work in the same county, creating shorter commute times
- Invest wisely, well-insured, comfortable banking online or in person
- News junkies (read a daily newspaper, watch news on TV, and go online for news)
- Enjoy reading, renting movies, playing board games and cards, doing crossword puzzles, going to museums and rock concerts, dining out, and walking for exercise

Segments: *5A Comfortable Empty Nesters; 5B In Style; 5C Parks and Rec; 5D Rustbelt Traditions; 5E Midlife Constants*

LifeMode 6 Cozy Country Living

- Empty nesters in bucolic settings
- Largest Tapestry group, almost half of households located in the Midwest
- Homeowners with pets, residing in single-family dwellings in rural areas; almost 30% have 3 or more vehicles and, therefore, auto loans
- Politically conservative and believe in the importance of buying American

- Own domestic trucks, motorcycles, and ATVs/UTVs
- Prefer to eat at home, shop at discount retail stores (especially Walmart), bank in person, and spend little time online
- Own every tool and piece of equipment imaginable to maintain their homes, vehicles, vegetable gardens, and lawns
- Listen to country music, watch auto racing on TV, and play the lottery; enjoy outdoor activities, such as fishing, hunting, camping, boating, and even bird watching

Segments: *6A Green Acres; 6B Salt of the Earth; 6C The Great Outdoors; 6D Prairie Living; 6E Rural Resort Dwellers; 6F Heartland Communities*

LifeMode 7 Ethnic Enclaves

- Established diversity—young, Hispanic homeowners with families
- Multilingual and multigenerational households feature children that represent second-, third- or fourth-generation Hispanic families
- Neighborhoods feature single-family, owner-occupied homes built at city's edge, primarily built after 1980
- Hard-working and optimistic, most residents aged 25 years or older have a high school diploma or some college education
- Shopping and leisure also focus on their children—baby and children's products from shoes to toys and games and trips to theme parks, water parks or the zoo
- Residents favor Hispanic programs on radio or television; children enjoy playing video games on personal computers, handheld or console devices
- Many households have dogs for domestic pets

Segments: *7A Up and Coming Families; 7B Urban Villages; 7C American Dreamers; 7D Barrios Urbanos; 7E Valley Growers; 7F Southwestern Families*

LifeMode 8 Middle Ground

- Lifestyles of thirtysomethings
- Millennials in the middle: single/married, renters/homeowners, middle class/working class
- Urban market mix of single-family, townhome, and multi-unit dwellings
- Majority of residents attended college or attained a college degree
- Householders have ditched their landlines for cell phones, which they use to listen to music (generally contemporary hits), read the news, and get the latest sports updates of their favorite teams
- Online all the time: use the Internet for entertainment (downloading music, watching YouTube, finding dates), social media (Facebook, Twitter, LinkedIn), search for employment
- Leisure includes night life (clubbing, movies), going to the beach, some travel and hiking

Segments: *8A City Lights; 8B Emerald City; 8C Bright Young Professionals; 8D Downtown Melting Pot; 8E Front Porches; 8F Old and Newcomers; 8G Hardscrabble Road*

LifeMode 9 Senior Styles

- Senior lifestyles reveal the effects of saving for retirement
- Households are commonly married empty nesters or singles living alone; homes are single-family (including seasonal getaways), retirement communities, or high-rise apartments
- More affluent seniors travel and relocate to warmer climates; less affluent, settled seniors are still working toward retirement
- Cell phones are popular, but so are landlines
- Many still prefer print to digital media: Avid readers of newspapers, to stay current
- Subscribe to cable television to watch channels like Fox News, CNN, and The Weather Channel
- Residents prefer vitamins to increase their mileage and a regular exercise regimen

Segments*: 9A Silver & Gold; 9B Golden Years; 9C The Elders; 9D Senior Escapes; 9E Retirement Communities; 9F Social Security Set*

LifeMode 10 Rustic Outposts

- Country life with older families in older homes
- Rustic Outposts depend on manufacturing, retail and healthcare, with pockets of mining and agricultural jobs
- Low labor force participation in skilled and service occupations
- Own affordable, older single-family or mobile homes; vehicle ownership, a must
- Residents live within their means, shop at discount stores and maintain their own vehicles (purchased used) and homes
- Outdoor enthusiasts, who grow their own vegetables, love their pets and enjoy hunting and fishing
- Technology is cost prohibitive and complicated. Pay bills in person, use the yellow pages, read newspapers, magazines, and mail-order books

Segments*: 10A Southern Satellites; 10B Rooted Rural; 10C Diners & Miners; 10D Down the Road; 10E Rural Bypasses*

LifeMode 11 Midtown Singles

- Millennials on the move—single, diverse, urban
- Millennials seeking affordable rents in apartment buildings
- Work in service and unskilled positions, usually close to home or public transportation
- Single parents depend on their paycheck to buy supplies for their very young children
- Midtown Singles embrace the Internet, for social networking and downloading content
- From music and movies to soaps and sports, radio and television fill their lives
- Brand savvy shoppers select budget friendly stores

Segments*: 11A City Strivers; 11B Young and Restless; 11C Metro Fusion; 11D Set to Impress; 11E City Commons*

LifeMode 12 Hometown

- Growing up and staying close to home; single householders

- Close knit urban communities of young singles (many with children)
- Owners of old, single-family houses, or renters in small multi-unit buildings
- Religion is the cornerstone of many of these communities
- Visit discount stores and clip coupons, frequently play the lottery at convenience stores
- Canned, packaged and frozen foods help to make ends meet
- Purchase used vehicles to get them to and from nearby jobs

Segments*: 12A Family Foundations; 12B Traditional Living; 12C Small Town Simplicity; 12D Modest Income Homes*

LifeMode 13 Next Wave

- Urban denizens, young, diverse, hard-working families
- Extremely diverse with a Hispanic majority, the highest among LifeMode groups
- A large share are foreign born and speak only their native language
- Young, or multigenerational, families with children are typical
- Most are renters in older multi-unit structures, built in the 1960s or earlier
- Hard-working with long commutes to jobs, often utilizing public transit to commute to work
- Spending reflects the youth of these consumers, focus on children (top market for children's apparel) and personal appearance
- Also a top market for movie goers (second only to college students) and fast food
- Partial to soccer and basketball

Segments*: 13A International Marketplace; 13B Las Casas; 13C NeWest Residents; 13D Fresh Ambitions; 13E High Rise Renters*

LifeMode 14 Scholars and Patriots

- College and military populations that share many traits due to the transitional nature of this LifeMode Group
- Highly mobile, recently moved to attend school or serve in military
- The youngest market group, with a majority in the 15 to 24 year old range
- Renters with roommates in nonfamily households
- For many, no vehicle is necessary as they live close to campus, military base or jobs
- Fast-growing group with most living in apartments
- Part-time jobs help to supplement active lifestyles
- Millennials are tethered to their phones and electronic devices, typically spending over 5 hours online every day tweeting, blogging, and consuming media
- Purchases aimed at fitness, fashion, technology and the necessities of moving
- Highly social, free time is spent enjoying music, being out with friends, seeing movies

Segments*: 14A Military Proximity; 14B College Towns; 14C Dorms to Diplomas;*

Source: ESRI (2020); https://doc.arcgis.com/en/esri-demographics/data/tapestry-segmentation.htm