A New Model for Spatial Analysis Site Selection and Decision Making for Small Retail Facilities: A Case Study for Starbucks- Seattle

Omar I. Aboulola
University of Jeddah

Strategic planners are often challenged by difficult spatial resource allocation decisions when analyzing a successful location for a new facility. This study demonstrated that the inclusion of additional social media activity variables (e.g., Twitter or Yelp) into a model significantly improved the analytic power regarding store sales, and hence, helped identify a successful store location. The regression analysis demonstrated that the increased R^2 value, resulting from the inclusion of the social media activity variables, improved upon the baseline model, and therefore helped to support strategic planners and decision makers regarding siting a new facility.

INTRODUCTION

Site selection is a critical aspect of strategic planning for a broad spectrum of public and private organizations (Owen and Daskin 1998). Strategic planners are often challenged by difficult spatial resource allocation decisions in determining the best locations for new facilities (Owen and Daskin 1998; Shelton and CEcD 2016). Site selection is thus an important strategic challenge. Therefore, it is important to identify the challenges and failures that have occurred previously to fully understand this problem.

Location selection is necessary in both public and private facilities especially when one considers how our environment has changed constantly (i.e., demographically, population, urban planning); therefore, understanding business needs and ensuring the business profits requires that spatial location selection research be obtained and accomplished (Marks et al. 1996). (Allahi et al. 2015) emphasized that location selection is very important because it could be very costly if new locations are not considered appropriately.

Location convenience is an important factor when customers select a financial institution (Kuo et al. 2002). For example, a customer may find a bank convenient if it has a branch or an Automated Teller Machine (ATM) near his/her residence or workplace. To stay competitive, retain existing customers, and acquire new ones, banks usually attempt to increase convenience by expanding their bank and/or ATM networks either by opening a new branch, new ATMs, or acquiring other ATMs. Expanding their ATM and/or branch network is usually done in two ways: building new facilities for branches and ATMs, or acquiring potential partner's locations. Moreover, it has been declared that the number of competitive locations has been an important factor for site selection, as competitiveness has been shown to affect market share in commercial areas (Chou et al. 2008).

Because location is such an important role in profit and revenue sustainability, poor spatial planning has caused many businesses to close or go bankrupt. Marros (2005) said that a company must determine its target market and the appropriate location that achieves the market target. Although that sounds simple and straightforward, it has proved difficult to consistently determine the most appropriate location to reach the market target, as is evidenced below.

Many well-known businesses are facing challenges to profit and revenue sustainability. For example, "In 2007 and 2008, Starbucks' CEO Howard Schultz was forced to come out of retirement to close hundreds of stores, and rethink the company's strategic growth plan" (Thau, Barbara 2014). Besides economic issues, location is another key factor in why specific branches experience an extreme decrease in sales ultimately causing them to close. Therefore, understanding how to optimally choose new locations with sustainable, increasing sales is businesses target (Thau, Barbara 2014).

As mentioned above, banks are vulnerable a well. In another example of large scale closures, (Egan 2015) reported that Bank of America's (BofA) footprint was shrinking:

"Two years ago, there were 5,328 U.S. branches at Bank of America. That has steadily declined every quarter since then, shrinking to 4,789 as of the end of the second quarter, a 10% drop. Fewer branches mean BofA also needs fewer employees. The bank's workforce has declined by nearly 16% over the past two years to 257,158 today. BofA-branded ATM machines are also being slowly sliced out. They have dipped by 2% over the last two years to nearly 16,000". (Egan 2015).

This many closures implies that there is an issue that caused business to close and staff to lose their jobs; it is certain that it is due to economic reasons, however location, location, location has significant impact on it and illustrates the importance of spatial location suitability (Egan 2015).

PROBLEM STATEMENT AND RESEARCH QUESTIONS

Site selection has often been based on a decision maker's experience or sophisticated spatial and mathematical models (Cheng et al. 2007). Site selection has been a problem for businesses for many years.

Despite the extended amount of time that people have been doing site selection, those models are underdeveloped and do not perform well. Furthermore, reports show that many businesses that use these models are closing(Egan, Matt 2015).

1. How can the development of a model ensure decision makers' and strategic planers' will receive an improved location prediction that considers' business, geographic, economic, and social media factors for site selection?
2. How can the addition of new factors (i.e. social media) will predict better sales than the current location factors'?

DECISION MAKING IN RETAIL SITE SELECTION

Decision making can have different and multiple meanings, but they all lead to the need to cognitively process information and to act among the possible alternative options that are available. A decision is the final act of selecting and deciding something. Decision making in site selection is the act to decide and select where to locate the potential site that increases the number of customers and impacts organizational profit not only in the selected location but also in other branches. Deciding where to locate a new potential location needs cognitive processing based upon the market surroundings and competitor challenges (Reynolds and Wood 2010).

In her review, Elisa Arrigo (2015) states that the selection of the best or optimum location has been considered by retailers as a strategic decision (Jiménez Capilla et al., 2016). In addition, her review shows that store locations are considered the most important determinant for retail businesses to succeed (Ghosh & Craig, 1983). Ghosh and Craig (1983) explained the reason store location is so important is because "it could provide strategic advantages that are difficult to overcome by competitors."

Clarke et al. (1997) framed it when they stated that the growing complexity of locational planning and the differences in decision-making processes made these early models "weak." When experts agree that the available models are weak, it is no wonder that planning is not easy and is a complicated task. Add the burden of knowing that the decision of the location needs to ensure optimum locations with revenue sustainability and it could be seemingly impossible. It is no wonder that "as many differences exist in practice across the different retail sectors, the location decisions ultimately rests on micro-scale considerations, that is the appropriateness or otherwise of the precise location within the chosen center" (S. Brown, 1994, p. 10).

A recent example for the use of GIS; the WAWA organization sought out Esri's help using their "Tapestry Data Product". (Berk 2016) explained that this software will help this retailer decide where to locate their new stores in Florida WAWA's first stores outside of the Northeast region. Another example of using GIS software is Tango Analytics that is used to analyze demographics, presence of competitors, and traffic trend (Berk 2016).

LITERATURE REVIEW

The study by Awaghade et al. (2014) is focusing on financial sectors and especially retail companies and banks, which grows and spreads continually. Cheng et al. (2007) presents its utility for shopping mall location selection, which is one of the core business activities of developers for long-term capital investment. Jankowski and Richard (1994) explains that urbanization is a very important continuous process to identify suitable housing areas for future development. Moreover, selecting the appropriate location for housing sites is a complex decision and process while there is not only technical requirement, but also physical, economical, social, environmental, and political requirement are included and important. Study (Ballou and Masters 1993) describes the factors that affect decision makers' in choosing the optimal location, which are inventory policy, facility location, and transportation and network. Study Awaghade et al. (2014) uses GIS and Remote Sensing (RS) to locate the potential sites for establishing Automated Teller Machines (ATM) centers in Aundh area of Pune city, India. Awaghade et al. (2014) focuses on showing where the existing ATMs are and where the suggested and suitable ATMs should have been or the areas that need ATMs' to fulfill the business needs and reach the banks goals (Awaghade et al. 2014).

On the other hand, another study Cheng et al. (2007) focused on showing how technology, specifically GIS applications with the use of maps can show information that can be understood and acknowledged to present spatial and non-spatial locations. Berman et al. (1992) showed two examples of a growing number of flexible service facilities that are ATMS' and gas stations. Jankowski and Richard (1994) indicated in another study for spatial location evacuation that spatial location and optimum decisions to allocate facilities and especially for safety and predicting disasters have been on the play since longtime and is so important for human safety. It was explained that deciding where to allocate fire stations, police departments, and evacuation disaster shelters is one of the most factors or reasons of GIS spatial location science. For instance, Florida is one of the most vulnerable coastal state to hurricanes. It is important to prompt people of an incoming hurricane and evacuate them.

Ballou and Masters (1993) explained the integrated approach of the warehouse site selection process where they relayed upon a theory of measurement for dealing with quantifiable and intangible criteria; this theory is called the analytic hierarchy process (AHP). Berman et al. (1992) Indicates that their outcome "system planner" focuses on determining the most accurate and best suitable facility locations is more concerned to place these facilities along paths of customers flow rather than placing them near the center of a cluster of residence or work places. Another study for Jankowski and Richard (1994) includes a model that helps to evaluate the possible locations for decision makers to build housing areas with factors such as population and poverty.

DATA

This study focused on Seattle due to its strong connection to social media. The availability of tweets data in Seattle helped us to proceed and test the validity of the study. This data has been categorized and retrieved from three different resources that are locational factors, demographic data, and social media data.

Starbucks, Seattle data was derived from the directory database at the Claremont Colleges. This database provides detailed information and statistics on over "13 million businesses, 120 million households" as well as detailed information on healthcare providers (Databases: Reference USA). They represent information for all three domains Education (schools, colleges), Health (clinics, urgent cares, and hospitals), and Businesses' (banks, malls, and other retail businesses such as Starbucks).

The focus of this paper is on Starbucks as a retail business. The database has showed that there are 466 Starbucks locations at Seattle. The attributes "factors" are Company Name, Address, City, State, ZIP Code, County Metro Area, Neighborhood, Primary SIC Description, Primary SIC Year Appeared, Latitude, Longitude, Location Employee Size Actual, Location Sales Volume Actual, and Square Footage.

Moreover, there are other key factors, which are the nearest five Starbucks to a specific Starbucks location (by using GIS tools to measure the distance), the nearest five Competitor to a Starbucks, the rating of each

In addition to the previous locational and social media factors, there are other important demographic factors that are crucial for site selections such as weather layer, Poverty layer, population "density" layer, gender rate layer, race rate layer, crime rate layer, education rate layer, and the age rate layer. All these are independent variables while the dependent variable is the actual sales of the location.

METHODOLOGY

This paper follows a design science research approach (DSR). Where DSR creates and evaluates IT artifacts to solve identified organizational problems (Hevner et al. 2004). DSR is an iterative research approach to design, build, and evaluate IT artifacts. Hevner and Chatterjee (2010) define DSR as "A research paradigm in which a designer answers questions relevant to human problems via the creation of innovative artifacts, thereby contributing new knowledge to the body of scientific evidence. The designed artifacts are both useful and fundamental in understanding the problem" (Hevner and Chatterjee 2010).

In addition, March and Smith (1995) have mentioned that there are two design processes and four design artifacts that have been produced by design-science research in IS. The two processes are build and evaluate. The artifacts are constructs, models, methods, and instantiations. Therefore, this study will build a model that predicts retail spatial location for decision makers and the model will be evaluated to ensure its significance and accuracy.

Using the selected factors with the available data leaded to assess the data to ensure it is cleaned and acceptable to be used in regression. In the process to create the artifact "model", which is the regression model, it was required to ensure the data is Cleaned, Formatted, Transformed non- normalized data.

Cleaned

Some fields were not important or included missing values and were therefore deleted or excluded. These fields were: ZIP_Code, ZIP_Four, Primary_SIC_Code, Primary_SIC_Description, IUSA_Number, FEATTYP, FT, F_JNCTID, T_JNCTID, METERS, FRC, NET2CLASS, NAMETYP, KPH, MINUTES, SPEEDCAT.

Formatted

Once a fully merged data set was prepared, it was uploaded to SPSS for more cleaning, coding, and modifying prior to analysis and modeling. The next section describes the process and steps implemented to recode the variables and transform non-normalized data for analysis and modeling.

Recoded Variables

To perform multiple regression, it was necessary to ensure all variable were numeric. By looking at the variable view in SPSS, it was clear that some variables were not numerical, while other variables were numerical. Specifically, two variables (road type and location size) needed to be recoded.

Road type had 16 different labels and could not be inserted as string or code from 1 to 16. To recode this variable, each road type (road, street, avenue, etc.) became a new dichotomous variable (0 or 1). Figure 2.9 shows the recoding for the road type variables.

Another variable that needed to be recoded was the square footage of each location, which was either 1500 – 2499 or 2500 – 4999. This data was recoded to 1 (1500 – 2499) and 2 (2500 – 4999) so that it could be analyzed. Figure 2.10 shows the recoding for the road type variables.

Transformed Non- normalized Data

After recoding the variables, it was then necessary to ensure that each variable was normally distributed to ensure a linear relationship between the variables. A statistical report for all variables was analyzed to show the normal distribution, mean, minimum, maximum of each variable. In addition, the skewness of each variable was considered to understand the distribution of each variable. The process showed which variables have a skewness that is out of the normal distribution range and which were confirmed to keep and use.

Skewness range has been a debatable subject. Brown (2008) emphasized different skewness ranges where some emphasize ± 1 while others emphasized ±2. Few researchers and statisticians emphasized ± 3. The skewness range relates to the field of study and the sample size of the data. Moreover, for this study the skewness range is between ± 3; in other words, regardless the sign, the number should not exceed 3 (Brown, 2008). The variables that are indicators and have skewness above 3, should be normalized.

Table 1, shows that the dependent variable has a skewness value 0.766. Of all 16 indicators, data showed that 14 variables (of the 19 variables) had skewness in an acceptable range that is between ± 3; while others had skewness that were out of the acceptable range ± 3.

TABLE 1
SKEWNESS FOR VARIABLES

Variable Name	Skewness	Normalized
Location_Sales_Volume	.766	Yes
Location_Size_Actual	.973	Yes
Count Tweets	10.367	No
Coding for location size	.287	Yes
road type is a Mall	3.121	No
road type is a Ave	.692	Yes
road type is a Rd	3.121	No
Distance Road	1.719	Yes
NEAR_DIST	2.233	Yes
AVG_10_SB2COMP	3.420	No
2014 Total Population Age 20-24	4.240	No
2014 Total Population Age 40-44	.894	Yes

2014 Asian Population	1.614	Yes
2014 Population Age 25+ GED/Alternative Credential	.623	Yes
2014 Population Age 25+: Bachelor's Degree	1.267	Yes
2014 Population Age 25+: Graduate/Professional Degree	1.577	Yes
2014 Household Income $150,000-$199,999	1.493	Yes
2014 Household Income $200,000 or greater	1.584	Yes
2014 Average Household Income	.910	Yes

The variables that were out of acceptable skewness range were: count Tweets (10.367), road type is a Mall (3.121), road type is a Rd (3.121), AVG_10_SB2COMP (3.420), and 2014 Total Population Age 20-24 (4.240). These five variables need to be normally distributed so that the variables can be used in analysis. Figure 2.12-2.14 are plot diagrams and histograms that show the distribution over the mean for the count Tweets, average distance of nearest competitor to Starbucks, and total population of age 20-24 indicators.

Correlate Variables

After recoding the needed fields to ensure they were numerical and cleaning the data to ensure that no fields were empty or null, a correlation analysis was performed to identify the relationship (correlation) between each IV and the DV. If there was no significant correlation between the IV and DV then these IVs needed to be excluded (in the regression). In addition, after determining the correlations between the dependent and independent; it was necessary to examine the relation between the independent correlated variables themselves; this is not a procedure for correlation, but is an examination that is checked before performing the regression. If there was a strong correlation between any two IVs, one of was excluded (in the regression). In this case, the Appendix C shows that there was a high correlation between graduate and the key indicators: income 150, income 200, and bachelor (degree earned). Also, there was a high correlation between average income with both income 150 and income 200. Therefore, it was necessary to exclude the key indicators that had more high correlation with the other key indicators such as bachelor, income 150, and income 200 in the regression. Afterwards, it was necessary to run the correlation again to ensure no further multicollinearity between these included indicators remained to do regression. After doing so, only the variables that remained were used in regression analysis when it was run.

Indicators (IVs)

Coding for location size, Count_Tweet_Log10, road type is a Ave, road type is a Rd, Distance_Road, NEAR_DIST, 2014 Average Household Income, 2014 Household Income $150,000-$199,999, 2014 Household Income $200,000 or greater, 2014 Asian Population, 2014 Population Age 25+: GED/Alternative Credential, 2014 Population Age 25+: Bachelor's Degree, and 2014 Population Age 25+: Graduate/Professional Degree are key indicators for the dependent variable (sales).

Model

Using the key indicators as IV with the DV (Sales) in a regression model, has shown which variables are significant and which are not. Then a regression model has been created as below:
Location_Sales_Volume = β + β * Coding for location size + β * Distance Road + β * 2014 Average Household Income + β* Count_Tweet_Log10 + β * 2014 Asian Population

Evaluation

This study will use SPSS to run a regression model. A pre-model and post-model evaluation will be conducted to show the effect of social media factor on the R^2 and adj R^2.

Pre-Model

The regression compiled the key predictors without including the Tweet count to acknowledge the coefficients data without the Tweets, so that it could later be compared to the post-model that did include the Tweets count. Table 4 (see appendix), show the weight of each predictor including the significance of these predictors. Table 2 (bellow) shows the R^2 that is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression indicates how well a regression model predicts responses for new observations. In other words, the model explains 27.3% of the response variable variation found the regression model. Moreover, the model explains 27.3% of the variability of the response data around its mean.

TABLE 2
PRE-MODEL SUMMARY

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.522[a]	0.273	0.266	331887.033
a. Predictors: (Constant), 2014 Average Household Income (Esri), Distance_Road, Coding for location size, 2014 Asian Population (Esri)				

The regression was conducted with the variables of location size, road-distance, House Hold Income, Asian population. The model explains 27% (R2) of the variance of the dependent variable.

Post-Model

Table 5 (see appendix) shows the significance of predictors from most significant to the least significant in order. Location size is the most significant, followed by the location's distance to road, then the household income level at a p-value of .06, the number of Tweets at a p-value of .013, and lastly, the least significant being Asian population with a p-value of .028. The standardized coefficients beta is the weight of each predictor which represents the importance of effect on the dependent variable "sales."

TABLE 3
POST- MODEL SUMMARY

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.552[a]	0.304	0.289	327039.395
a. Predictors: (Constant), 2014 Population Age 25+: Graduate/Professional Degree (Esri), Coding for location size, road type is a Ave, NEAR_DIST, Distance_Road, road type is a Rd , 2014 Population Age 25+: GED/Alternative Credential (Esri), Count_Tweet_Log10, 2014 Average Household Income (Esri), 2014 Asian Population (Esri)				
b. Dependent Variable: LocationSales_Volume				

In the post model, the regression was conducted with the variables of location size, road-distance, House Hold Income, Asian population, and Tweet-Count. The model explains 30% (R^2) of the variance of the dependent variable. By adding social media (Tweets) the model improved in explaining the variance of the dependent variable. This percent means that the model explains 30.4% of the variance to predict the sales (DV). Adding the Tweet accounts increased the R^2 of the pre-survey from 27.3% to 30.4 %, which indicates an effect. Tweets can play a role indicating where to predict future locations in addition to other factors that were considered besides high Asian populations, average household income, the location's size, and the location's distance to nearest road

Contribution

This study will contribute in defining a model that predicts and allocates new locations with showing the expected sales for multiple suggested location based on the availability of new factors that were trained, tested and validated upon current locations. The factors of social media are the key contribution while few or no studies have included them. As important as economics is for businesses, population density plays an important factor to show where consumers' and customers are. While crowds of people can serve as a general indicator, social media can provide their locations as well as their interest. Therefore, by knowing the count of the tweets of people everywhere, an assumption of the suggested place can have an affect based on the tweets as one of the key factors but not the only or most important factor. Overall, social media is a powerhouse and largely overlooked factor in spatial location planning. This study illustrates the value of this concept.

CONCLUSION

Optimum site selection has been considered by retailers and academic researchers to be a strategic decision for businesses. Therefore, past decades have showed a steep growth in the use of site selection (Jiménez Capilla et al. 2016).

For decades, researchers have been designing techniques to predict the optimal method for site selection. This literature review, explains the different techniques including the definition and importance of each and where it has been manipulated. Some of these techniques have been classified as mathematical while others were conceptual. In addition, the complexity of some of the deductive techniques (Huff Model, Gravity Model, checklists/analogue/ratio, and Multi Criteria Decision Making Model, etc.) has encouraged researchers to seek inductive approaches (Neural Network, Decision Trees, etc.) using machine learning techniques (Higgins et al. 2014; Reynolds and Wood 2010). Using these different methods with the selection of dependent and independent factors have contributed on selecting optimum location. These factors play an important role base on the type of business and the domain of the location. Previous studies show that demographics, transportation & road type, sales are more applied and increasable effects on sustainable revenue.

Researchers, who are adopting the techniques, have identified different factors for site selection. These factors have been classified as locational attributes (sales of current locations, trade area development, market potential, size of location, etc.) and demographical attributes (population, density, poverty, income, education, race, etc.). Nevertheless, very few of them have attempted to include social media as a factor to support site selection decision making. This paper examines the efficacy of prediction upon applying social media factor while site selecting retail locations. This paper outcome will inspire researchers, decision makers, and strategic planners to understand the different domains, methods, and history of retail site selection decision making with acknowledging the effect of social media.

REFERENCES

Allahi, S., Mobin, M., Vafadarnikjoo, A., & Salmon, C. (2015). *An Integrated AHP-GIS-MCLP Method to Locate Bank Branches*. https://works.bepress.com/amin_vafadarnikjoo/10/

Arrigo, E. (2015). The Role of the Flagship Store Location in Luxury Branding. An International Exploratory Study. *International Journal of Retail & Distribution Management*, 436), 518–537. https://doi.org/10.1108/IJRDM-08-2013-0158

Awaghade, S., Dandekar, P., & Ranade, P. (2014). SITE SELECTION AND CLOSEST FACILITY ANALYSIS FOR AUTOMATED TELLER MACHINE (ATM) CENTERS: CASE STUDY FOR AUNDH (PUNE), INDIA. *INTERNATIONAL JOURNAL OF ADVANCEMENT IN REMOTE SENSING, GIS AND GEOGRAPHY,* 2(1), 19–29. http://www.irosss.org/ojs/index.php/IJARSGG/article/view/255

Ballou, R. H., & Masters, J. M. (1993). Commercial Software for Locating Warehouses and Other Facilities. *Journal of Business Logistics*, 14(2), 71–107. http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=9707111016&site=ehost-live&scope=site

Berk, B. (2016). Wawa's Secret Weapon in Florida Expansion. *Convenience Store News*. http://www.csnews.com/node/90268, accessed February 21, 2017

Berman, O., Larson, R. C., & Fouska, N. (1992). Optimal Location of Discretionary Service Facilities. *Transportation Science*, 26(3), 201–211. http://pubsonline.informs.org/doi/abs/10.1287/trsc.26.3.201

Brown, S. (1994). *Retail Location at the Micro-Scale Inventory and Prospect.pdf.*

Cheng, E. W. L., Li, H., & Yu, L. (2007). A GIS Approach to Shopping Mall Location Selection. *Building and Environment*, 42(2), 884–892. https://doi.org/10.1016/j.buildenv.2005.10.010

Chou, T.-Y., Hsu, C.-L., & Chen, M.-C. (2008). A Fuzzy Multi-Criteria Decision Model for International Tourist Hotels Location Selection. *International Journal of Hospitality Management*, 27(2), 293–301. https://doi.org/10.1016/j.ijhm.2007.07.029

Clark, J. R., Matheny, N. P., Cross, G., & Wake, V. (1997). A Model of Urban Forest Sustainability. *Journal of Arboriculture*, 23, 17–30. https://www.researchgate.net/profile/Nelda_Matheny/publication/254202799_A_MODEL_OF_URBAN_FOREST_SUSTAINABILITY/links/544133820cf2e6f0c0f6040f.pdf

Columnist, J. T. (2015, December 2). Saying Goodbye to Beloved Ralph's Grocery. *The Seattle Times*. http://www.seattletimes.com/business/economy/saying-goodbye-to-beloved-ralphs-grocery/, accessed May 13, 2016

Databases: Reference Sources - Claremont Colleges Library. (n.d.). http://libraries.claremont.edu/resources/databases/bysubject.asp?SubjId=79, accessed November 13, 2015

Egan, M. (2015). Hundreds of Bank of America Branches Are Disappearing. *CNN Money*. http://money.cnn.com/2015/07/15/investing/bank-of-america-branches-layoffs/index.html, accessed May 13, 2016

Egan, M. (2015). Hundreds of Bank of America Branches Are Disappearing - Jul. 15, 2015. http://money.cnn.com/2015/07/15/investing/bank-of-america-branches-layoffs/, accessed April 4, 2016

Ghosh, A., & Craig, C. S. (1983). Formulating Retail Location Strategy in a Changing Environment. *Journal of Marketing*, 47(3), 56–68. https://doi.org/10.2307/1251197

Hevner, A., & Chatterjee, S. (2010a). *Design Science Research in Information Systems," in Design Research in Information Systems.* Integrated Series in Information Systems: Springer US. 9–22. http://link.springer.com/chapter/10.1007/978-1-4419-5653-8_2

Hevner, A., & Chatterjee, S. (2010b). *Design Science Research in Information Systems," in Design Research in Information Systems (Vol. 22).* Boston, MA: Springer US, 9–22. (http://link.springer.com/10.1007/978-1-4419-5653-8_2).

Hevner, A. R., March, S. T., & Ram, S. (2004). DESIGN SCIENCE IN INFORMATION SYSTEMS RESEARCH. *MIS Quarterly*, 28(1), 75–105.

Jankowski, P., & Richard, L. (1994). Integration of GIS-Based Suitability Analysis and Multicriteria Evaluation in a Spatial Decision Support System for Route Selection. *Environment and Planning B*, 21, 323–323.

Jiménez Capilla, J. A., Carrión, J. A., & Alameda-Hernandez, E. (2016). Optimal Site Selection for Upper Reservoirs in Pump-Back Systems, Using Geographical Information Systems and Multicriteria Analysis. *Renewable Energy*, 86, 429–440. https://doi.org/10.1016/j.renene.2015.08.035

Kuo, R. J., Chi, S. C., & Kao, S. S. (2002). A Decision Support System for Selecting Convenience Store Location through Integration of Fuzzy AHP and Artificial Neural Network. *Computers in Industry*, 47(2), 199–214. https://doi.org/10.1016/S0166-3615(01)00147-6

Linn, A. (2008). *Starbucks to Close 600 Stores in the U.S.* (http://www.nbcnews.com/id/25482250/ns/business-us_business/t/starbucks-close-stores-us/, accessed May 13, 2016).

March, S. T., & Smith, G. F. (1995). Design and Natural Science Research on Information Technology. *Decision Support Systems*, 15(4), 251–266. https://doi.org/10.1016/0167-9236(94)00041-2

Marks, D., Liebman, J., & Revelle, C. (1996). An Analysis of Private and Public Sector Location Models. *Location Theory*, 2, p. 405.

Marros, R. J. (2005). *Using a Geographic Information System (GIS) to Visualize and Analyze Spatial Location in a Retail Environment*. University of Minnesota, Winona. (http://www.gis.smumn.edu/GradProjects/MarrosR.pdf).

Mishra, S. (2009). GIS in Indian Retail Industry-a Strategic Tool. *International Journal of Marketing Studies*, 1(1), p. 50. http://www.ccsenet.org/journal/index.php/ijms/article/view/1647

Overmars, K. P., Groot, W. T. de, & Huigen, M. G. A. (2007). Comparing Inductive and Deductive Modeling of Land Use Decisions: Principles, a Model and an Illustration from the Philippines. *Human Ecology*, 35(4), 439–452. https://doi.org/10.1007/s10745-006-9101-6

Owen, S. H., & Daskin, M. S. (1998). Strategic Facility Location: A Review. *European Journal of Operational Research*, 111(3), 423–447. (http://www.sciencedirect.com/science/article/pii/S0377221798001866).

Peterson, H. (2015). Wal-Mart Suddenly Closed 5 Stores and Laid off Thousands of Workers and No One Knows Why. *Business Insider*. http://www.businessinsider.com/wal-mart-suddenly-closes-stores-2015-4, accessed May 13, 2016

Reagan, K. G., & Courtney. (2016). Wal-Mart to Close 269 Stores Globally as It Retools Fleet. *CNBC*, January 15. (http://www.cnbc.com/2016/01/15/wal-mart-to-close-269-stores-as-it-retools-fleet.html, accessed May 13, 2016).

Reynolds, J., & Wood, S. (2010). Location Decision Making in Retail Firms: Evolution and Challenge. *International Journal of Retail & Distribution Management*, 38(11/12), 828–845. https://doi.org/10.1108/09590551011085939
Shelton, B., & CEcD, H. (2016). *The Retail Site Selection Process*. https://www.buxtonco.com/pages/the-retail-site-selection-process/, accessed September 8, 2016
Starfish. (2015). *Starbucks: Brewing up a Data Storm! – Digital Innovation and Transformation*. (https://digit.hbs.org/submission/starbucks-brewing-up-a-data-storm/, accessed February 21, 2017).
Thau, B. (2014). How Big Data Helps Chains Like Starbucks Pick Store Locations : A Key to Retail Success. *Content Loop*. http://www.content-loop.com/big-data-helps-chains-like-starbucks-pick-store-locations-unsung-key-retail-success/
Yang, J., & Lee, H. (1997). An AHP Decision Model for Facility Location Selection. *Facilities*, 15(9/10), 241–254. (https://doi.org/10.1108/02632779710178785).

APPENDIX

TABLE 4
PRE-MODEL COEFFECIENTS

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	292750.047	68967.969		4.245	0.000		
	Coding for location size	360481.603	31572.118	0.461	11.418	0.000	0.968	1.033
	Distance_Road	-1590.129	300.471	-0.211	-5.292	0.000	0.991	1.009
	2014 Average Household Income (Esri)	1.354	0.618	0.090	2.192	0.029	0.943	1.060
	2014 Asian Population (Esri)	8.810	4.761	0.077	1.850	0.065	0.922	1.084

TABLE 5
POST-MODEL COEFFECIENTS

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	117509.609	118972.859		0.988	0.324		
	Coding for location size	346718.375	31709.332	0.443	10.934	0.000	0.935	1.070
	Distance_Road	-1282.303	315.318	-0.169	-4.067	0.000	0.884	1.131
	2014 Average Household Income (Esri)	2.208	0.792	0.146	2.786	0.006	0.560	1.784
	Count_Tweet_Log10	69882.254	27875.744	0.119	2.507	0.013	0.679	1.473
	2014 Asian Population (Esri)	13.526	6.147	0.117	2.200	0.028	0.539	1.856