Locating Affordable Grocery Business in Philadelphia

Xiaoran Wang
LARP 745 | MUSA 800 | Project Report

Abstract

Site selection is usually one of the crucial processes in marketing planning for any business. A well-suitable site could directly affect branding, customer flows, and long-term profits. Especially in an urban context where the market is relatively saturated, site selection is even more significant to the success of the new store. In recent decades, Geographic Information Science (GIS) has been widely used in this decision-making process of market planning. This project is going to use a GIS-based multi-criteria analysis to identify a potential site for a new affordable grocery store in the city of Philadelphia.

1. Introduction

Geography plays a crucial role in the success of a business. Opening a new store at a new site always requires complicated analysis to reduce the potential risks in branding and financial sides. Researchers have stated that geographical information systems (GIS) are able to provide valuable information and visualize the spatial characteristics to the business operators and to help them to make better decisions on site selection.

Philadelphia, as the fourth largest city in the United States, is considered a mature market in the grocery retail business. Functional marketing evaluation of local grocery retail market in Philadelphia usually faces the challenges from the extreme class differentiation and uneven population distribution. The grocery retail market at such a historical city has already saturated, so from a pure marketing perspective, finding a suitable location for the new grocery store is also very problematic.

Based on the above facts of difficulty, GIS can be a useful tool to overlay different features and provide a more definite sense of both the socioeconomic structure and environmental characteristics of the city. This project will first review the past cases of GIS in the store allocation and then using the multi-criteria analysis from different levels to explore to the study area. Finally, a suitable location for the new grocery store will be provided based on the results of the investigation.

2. Objectives

The project aims to conduct the process of site selection for a large-scale affordable grocery store at Philadelphia. Kroger is the represented example of store in this project. The ideal establishments of the new grocery store potentially include the flowing factors:

a) 100,000 to 145,000 sq. ft large marketplace.
b) available parking space.
c) affordable pricing for fresh grocery.

The project will finish the primary and secondary multi-criteria analysis and select the location for the new grocery marketplace.

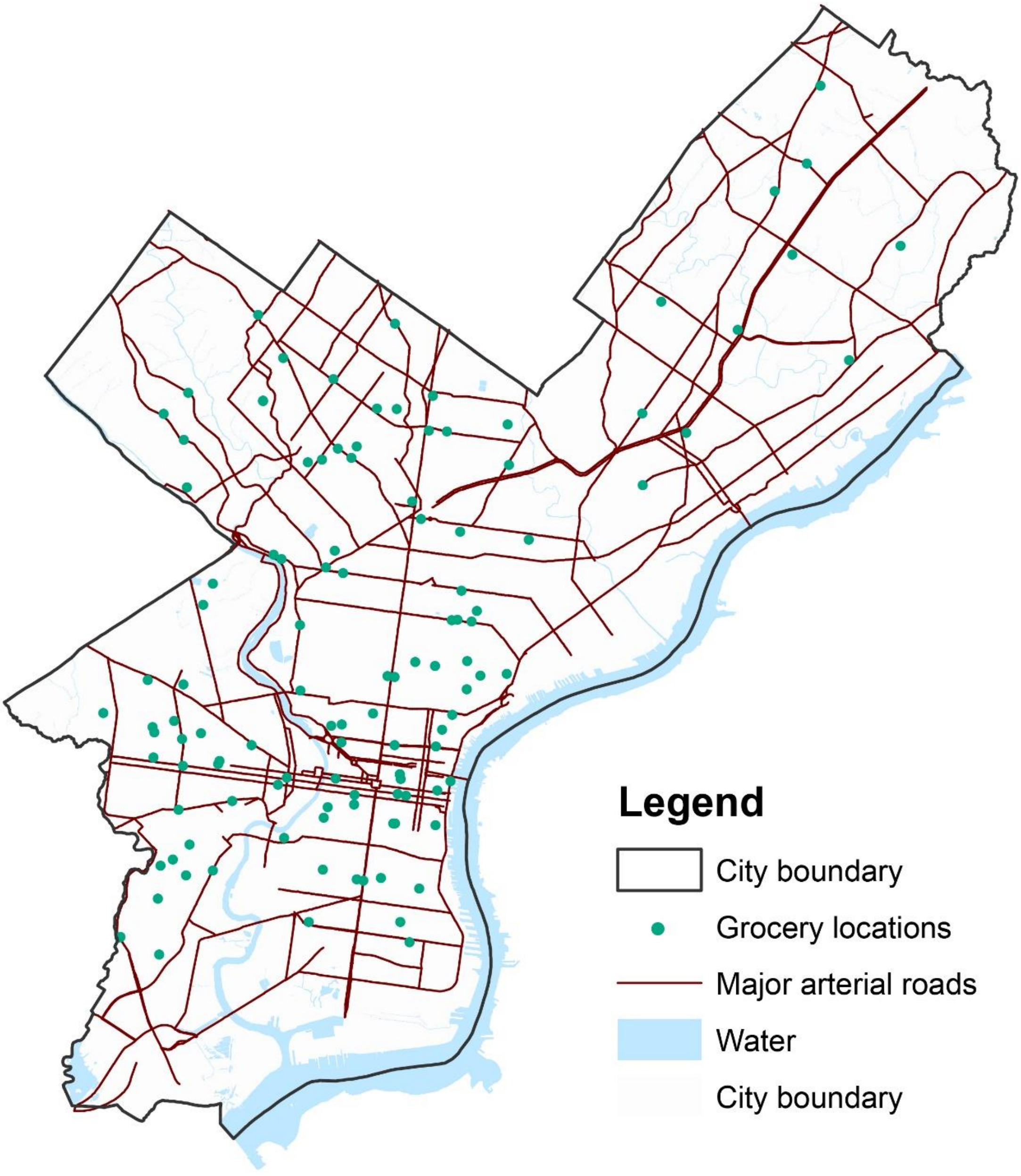


Fig.1 Map of Philadelphia grocery distribution

3. Literature Review

In the past, the traditional methods of industrial site selection were based on purely economic and technical criteria, which potentially include environmental factors, government regulations, as well as social-level demands. Thus, it was usually a typical multi-criteria analysis (Rikalovic., 2013). In that case, GIS is a powerful tool in spatial-level analysis which provides good sense of capturing, storing, analyzing and displaying geographic information to satisfy the multi-criteria and multi-objective location analysis (Rikalovic., 2013). It indicates that industrial site selection is a spatial problem, so map is a direct way to convey the distribution of multiple factors when doing the site selection. The researchers also believe that GIS would have extensive usage in the exploration of better multi-criteria decision methods in the future (Rikalovic., 2013).

Among all different kinds of market planning, retail planning primarily bases on issues like products, promotion, and location. These issues mainly rely on the customer's distribution and the location of competitors (Murad, 2003). Thus, before using GIS to do the planning, it is necessary to identify the nature of retail planning and fully understand the market environment. For the case of grocery store planning, the Miracle supermarket in Ontario, Canada, has made a great achievement. They used GIS as the major technique and firstly conducted a customer survey to existing stores to thoroughly investigate customer expectations and features. In addition, many projects mention the interaction model in deciding on site selection. Murad's project used the degree of interaction between existing stores and neighborhoods to evaluate how well the neighborhoods are served by the existing stores. He tended to build the new store at the regions are poorly served. Similarly, in Reilly's law of retail gravitation, the gravity model has also been applied to locating individual stores since it measures the degree of interaction between origins and destinations. Opposite to previous interpretation from Murad, this approach is from the perspective of the new store instead of existing stores. The paper discusses that the gravity model is also useful to estimate how many customers will be attracted by the new store and do not visit the previous existing stores. Still, in this case, one needs to repeat the full gravity model to all current stores, which is time-consuming and not very feasible (Eck & Jong, 1999).

The previous site selection projects always considered the established requirements, location accessibility, demographics, and competitions combined with being the core structure on evaluating the new sites (Roig-Tierno et al., 2013). The specific variables within each category would be assigned an index based on the type of retail business. Then, by overlaying data and maps and adding up all index, the potential sites would be decided by the rank of the total score. Then, in terms of unit size when doing the evaluation, the concepts of geodemand and geocompetition has been proposed to make the decision on site selection for a retail location. The data tends to be finer and more detailed, so the best unit usually downs to the city block level (Roig-Tierno et al., 2013). It would be necessary to give a proper estimation of the average number of residents and housing units in each census block. The methods would be more helpful on the secondary analysis to decide the micro-level candidate sites.

In this project, it would be hard to do the part of the customer survey so that the related information will be estimated from the census data and summarized by the Tapestry Segment from Esri. The project will follow the top-down skeleton of conducting the multi-criteria analysis with four core elements and measure the accessibility and interactions between the potential store locations and other geographical units in the study area.

4. Methodology

4.1 Data

Data used in this project are primarily collected from open source databases, such as OpenDataPhilly, U.S. Census Bureau, Tiger Census, OpenStreetMap, and Mapbox API.

Source	Data
OpenDataPhilly	Philadelphia boundary and water City roads Land use
U.S. Census Bureau	Block group level Population count (2013-2018)
Tiger Census	Block group and block shapefiles
OpenStreeMap API	Supermarkets and farmer's market locations
Mapbox API	Isochrones for different transportation modes

4.2 Framework

The project will apply a GIS-based multi-criteria analysis, and the general frame will include two major parts: macro-level analysis and micro-level analysis, so it follows a top-down workflow. The logistic behind the method is to gradually narrow down the potential area based on multiple factors and choose the final selected location(s). Different geospatial units will be applied to each level: block group for macro-level, and block for micro-level.

5. Analysis

5.1 Macro-level analysis

At this level, the goal is to select a certain number of potential areas based on decision factors to narrow down the physical range. The physical characteristics are the most fundamental prerequisite, and the related researches have shown that one needs to consider both demand- and supply-side factors at the beginning stage of site screening. In such a way, I summarized that the road proximity, market competition, and population increasing trends are the three decision factors that correspond to the three perspectives in site screening, and I will examine these factors at this level of analysis.

Decision factor 1: Road proximity

The grocery location in the urban area should have higher accessibility, and the road is the key to measure it. I set 1km (0.62 miles) buffers around the major arterials roads to ensure accessibility to public transit and the convenience of driving-oriented customers. Since our study area is in an urban environment, where the infrastructures are well-constructed, and the terrain is usually suitable for developments, I do not include the physical factors such as slope and farmland in this project.

I also want to avoid that the costumers of the new store to be highly spatial overlapped with the customers of the existing business. Therefore, I draw a 0.25mile buffer around each current grocery locations and erase these grocery location buffers from the road buffers. This buffer radius refers to the acceptable walking distance in U.S. research studies (Yang & Diez-Roux, 2012). In terms of the driving-oriented customer, the average driving time to the grocery store for urban residents is approximately 15 minutes (Hamrick &Hopkins, 2012). However, according to the map of isochrones, the 15-min driving distance buffers of existing stores have already covered the study area, which makes the driving distance buffer less valuable in this case.

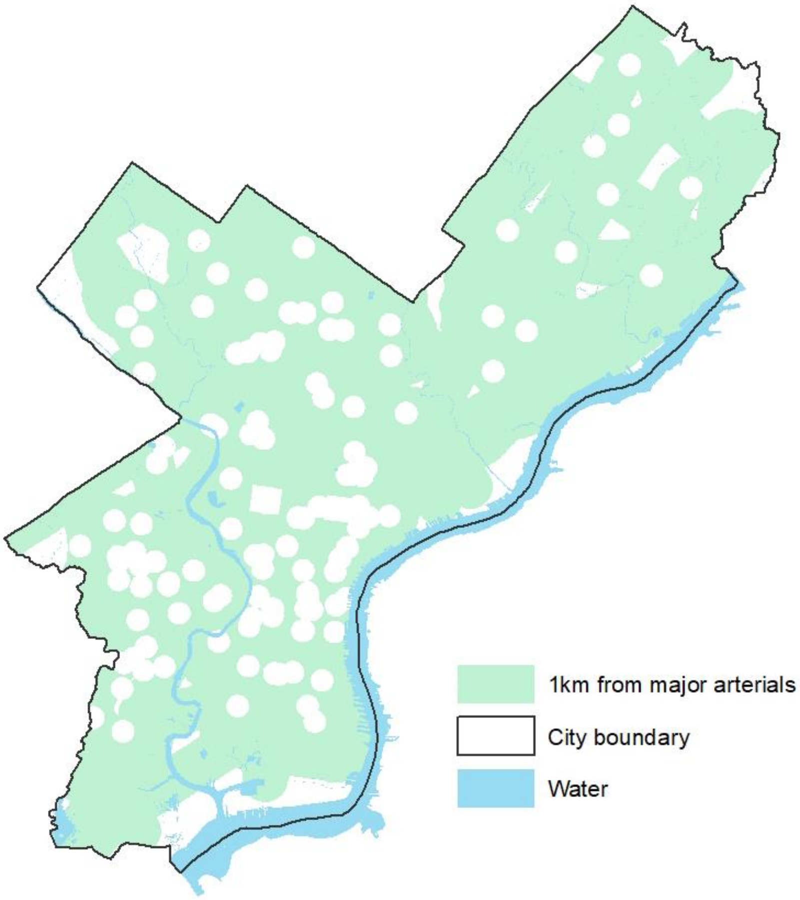


Fig. 2 Road proximity analysis: distance to major arterial road < 1km

Decision factor 2: Geo-competition

On the supply side, competition between similar grocery business is lack of positive impacts on promoting new store due to resident's grocery shopping habits. The area with a lower degree of competition would be a more suitable choice for the new store to develop business. Meanwhile, these areas usually accompanied by larger open spaces and lower land prices. Geo-competition analysis is required to fulfill the above supply-side requirements. The kernel density analysis is able to identify the areas with higher concentrations of existing grocery locations. With this method, space is divided into pixels with a standard unit to provide a smooth density map like Figure 3. Figure 4 categorizes three levels of geo-competition is resulting from a further reclassification of the previous continuous density map.

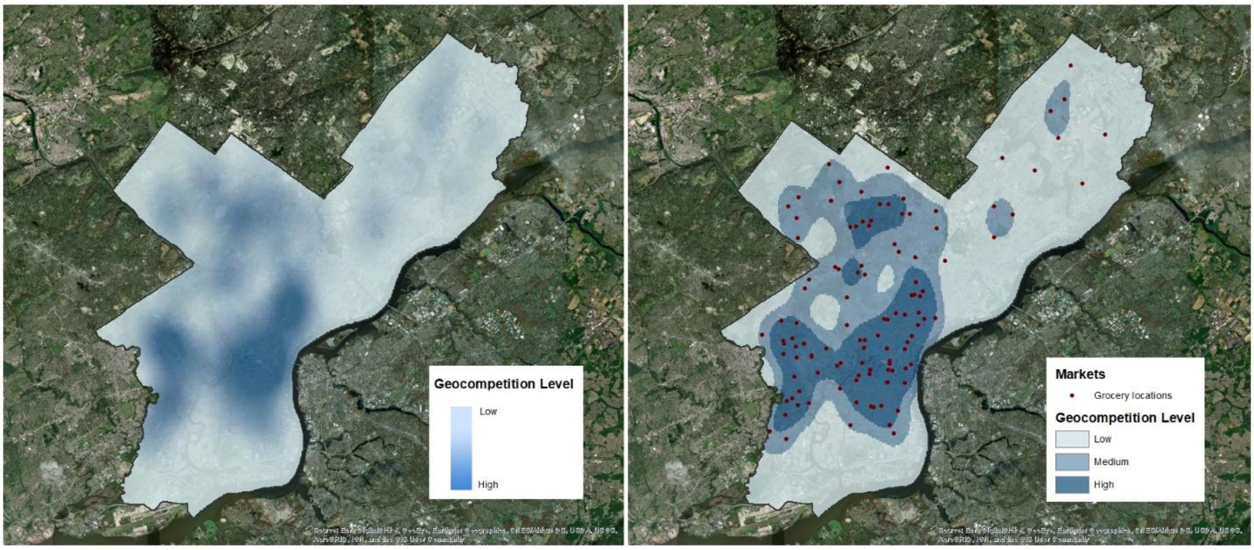


Fig. 3, 4 Geo-competition kernel density map and categorized map

Decision factor 3: Population growth pattern

Locating at a continuously growing neighborhood can benefit the new store from a long-run perspective. The customer base is a significant feature of the demand-side consideration for a new store. I use the block-group level population count data from 2013 to 2018 to inspect the population growth across the study area. The average population growth rate over the six years is calculated for each block group, respectively. To reflect a continuously developed pattern of the local population, the growth rate is a more interpretable feature than population count or density. Figure 5 shows the distribution of the population growth pattern.

We know that many block groups already have great population bases, so the population growth rates in these areas are not very high in the past six years compared to the block groups with small population bases. However, for metropolitans like Philadelphia, the local markets at the large population bases area have already saturated. The maps in at the geo-competition section have also proved it: higher geo-competition areas largely overlap the areas like the central city and west Philly.

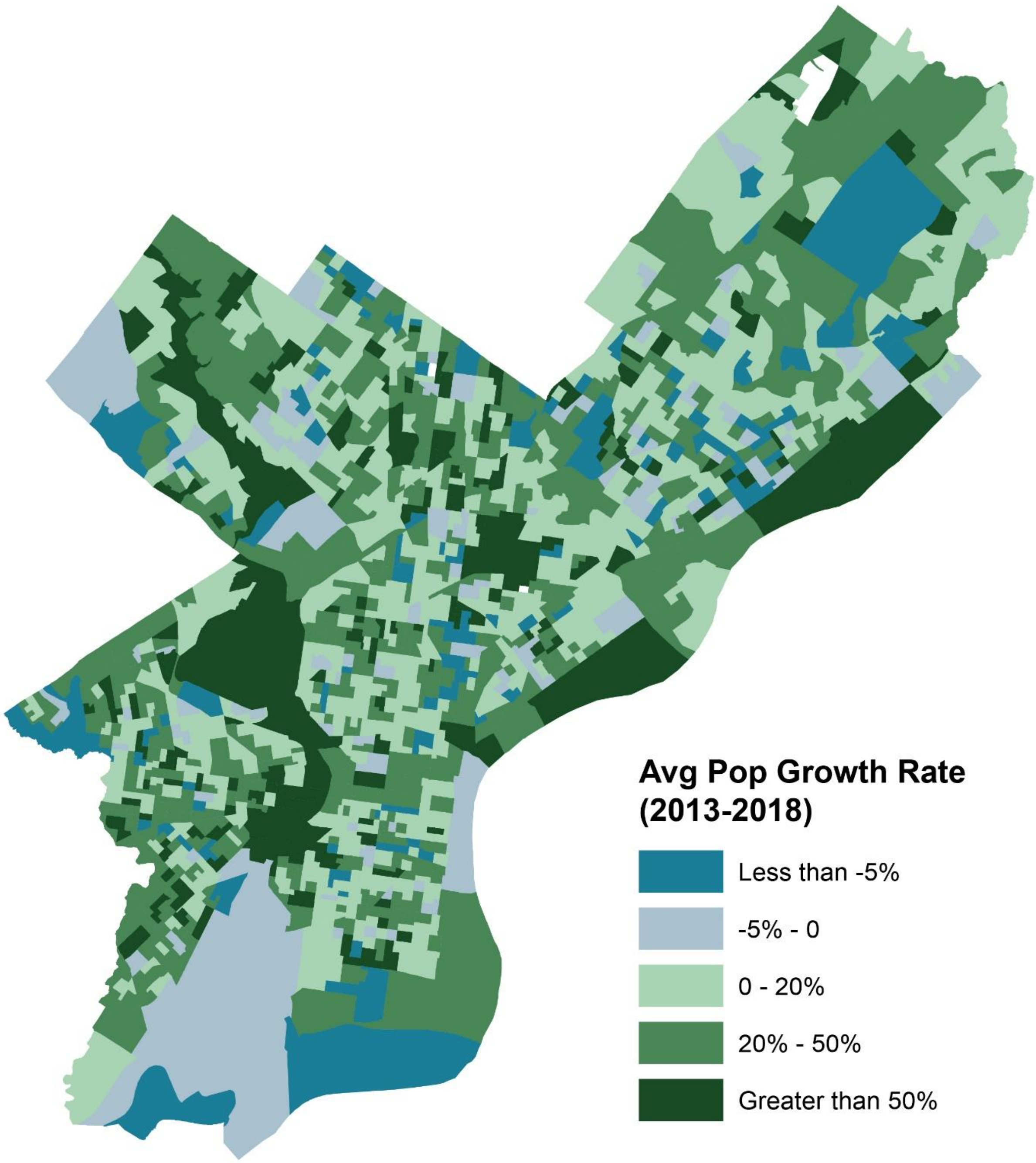


Fig. 5 Average population growth rate per block group (2013-2018)

Scoring potential areas

Prior research has categorized the study area based on three decision factors, and these criteria can pre-identify the potential area for the location of the new store. I use "Polygon to raster" to split the study area into more than 27000 pieces of the land segment, and each segment is with an area of 144,415 sq. ft. The polygon-to-raster conversion is also applied to transfer the vector shapefiles (road buffers and population growth rate) into raster layers. In the raster context, each decision factor layer is reclassified and assigned a weight according to the level of significance. After the reclassification, I use the raster calculator to add the weights of three layers of decision factors for each piece of the land segment to calculate the suitable score. The most suitable land segment would have a score of 10 after the calculation.

Decision factors		Weight
Road proximity	< 1km from major arterials	0
	> 1km from major arterials	3
Geo-competition	High	0
	Medium	2
	Low	3
Population growth rate	< -5%	0
	-5% - 0	1
	0 - 20%	2
	20% - 50%	3
	> 50%	4
		Highest suitable score: 10

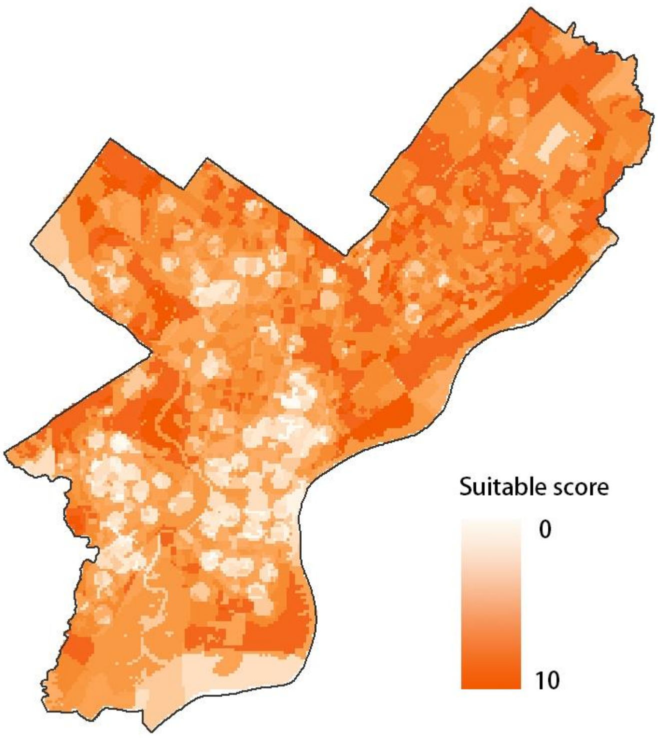
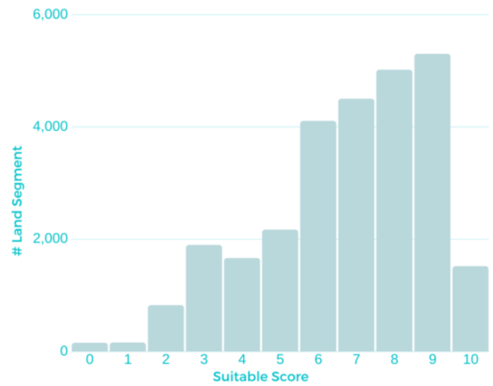
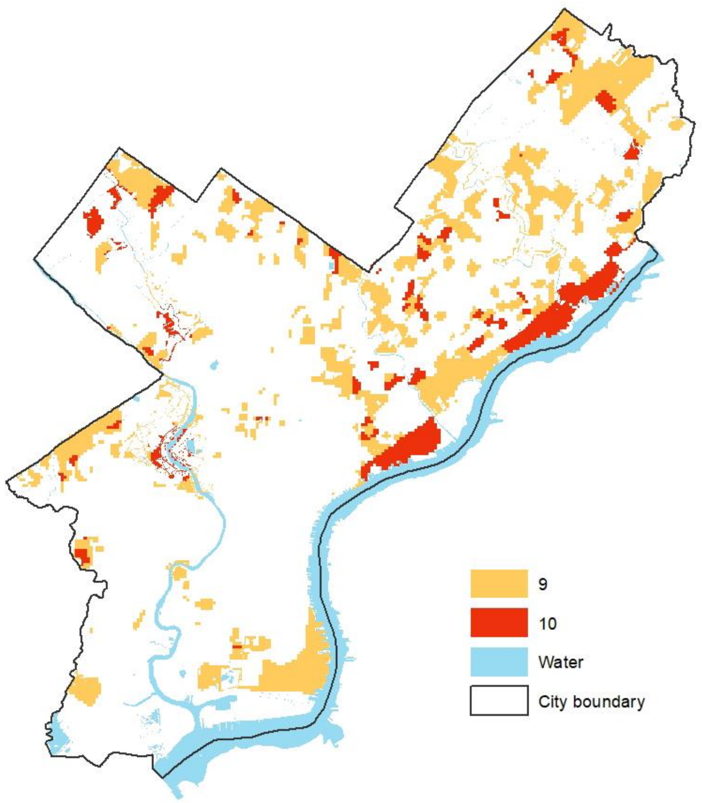


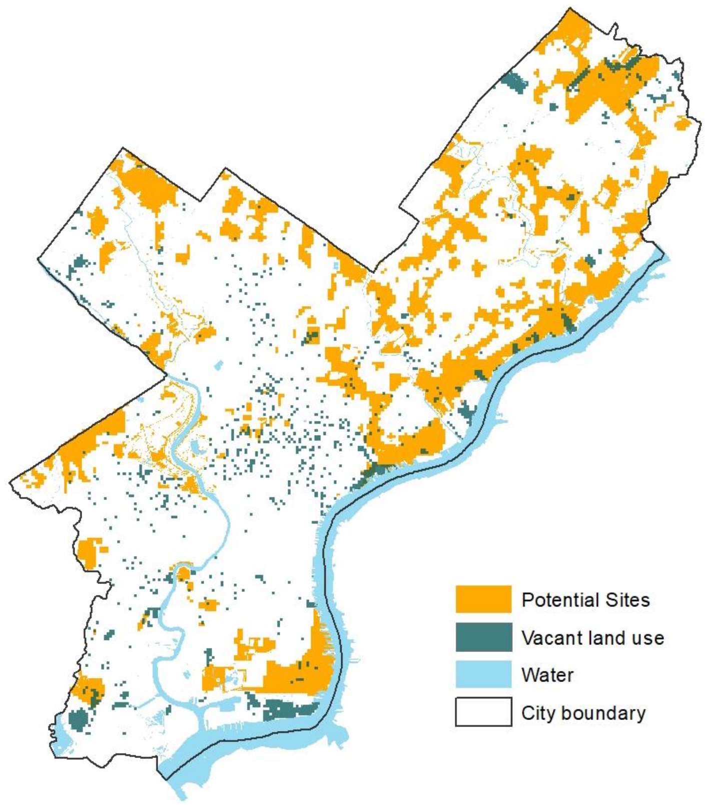
Fig. 6
The site suitability raster map demonstrates the score of land segment. The bar chat below shows the segment counts at each score.





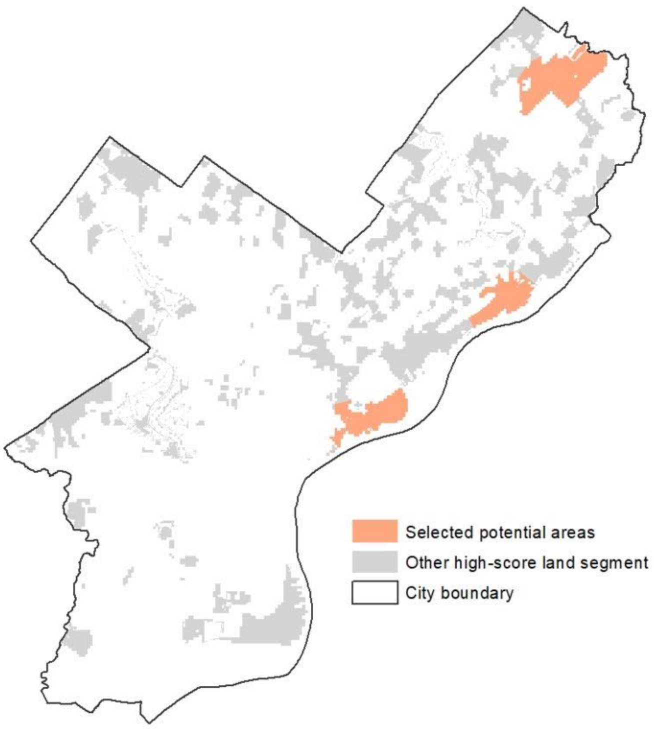
I found that some of the areas with higher scores are greenspaces, where should be excluded from the potential areas. Thus, I overlay the land use layer and erase the park and recreation facilities from the map. There are 5503 pieces of land segments classified as a potential site, which occupied 20% of the study area.

Fig. 7
The map only highlights the land segments with an above 9 suitable score, and the green spaces have been excluded.



At the last step of the macro-level analysis, I add a layer of vacant spaces since it provides a reference about which land segment contains more vacant space,

Fig. 8
The land segment pattern after adding the vacant space layer.



I finally use both satellite images and the vacant spaces layer to inspect the map of high suitable score land segments and decide three macro-level potential areas for conducting further analysis.

Fig. 9
Outcomes of the macro-level analysis. Orange pieces are the selected potential areas since they contain more vacant spaces, and the satellite images also show that the landscape in these areas has higher constructed availability.

5.2 Micro-level analysis

The micro-level analysis is a population-based evaluation for the three potential areas selected from the macro-level analysis. As mentioned in the previous section, the spatial unit at this stage is down to the block level, which allows a more detailed evaluation at the small scale of the area. According to the further observation of the combination of satellite image and vacant space layer, I identify one block per potential area with an average available land around 390,000 sq. ft, and this size is enough to cover both marketplace and free parking lot.

To have a better estimation of the customer base, the most up-to-date population data is necessary, but the U.S. census only has the block-level population count in 2010. Thus, I split the 2018 population in each block group and proportionally allocate the people in the blocks based on the block area. This method assumes that people are evenly distributed across space. The formula and examples are presented below:

GEOID	NAME	pop2018	BGarea
421010316004	Block Group 4, Census Tract 316, Philadelphia County, Pennsylvania	1517	1224154.59059

BLOCKCE10	NAME	Area	pop2018	estPop
4007	Block Group 4, Census Tract 316, Philadelphia County, Pennsylvania	55267.608831	1517	68
4006	Block Group 4, Census Tract 316, Philadelphia County, Pennsylvania	79650.320627	1517	99
4001	Block Group 4, Census Tract 316, Philadelphia County, Pennsylvania	96881.480834	1517	120
4000	Block Group 4, Census Tract 316, Philadelphia County, Pennsylvania	106547.637349	1517	132
4002	Block Group 4, Census Tract 316, Philadelphia County, Pennsylvania	112557.595618	1517	139
4003	Block Group 4, Census Tract 316, Philadelphia County, Pennsylvania	144371.05951	1517	179
4004	Block Group 4, Census Tract 316, Philadelphia County, Pennsylvania	313898.451131	1517	389
4005	Block Group 4, Census Tract 316, Philadelphia County, Pennsylvania	314980.43668	1517	390

$$\frac{A_{block}}{A_{blok\ group}} = \frac{P_{block}}{P_{block\ group}}$$

I draw a 2-mile buffer around each target site to cover the surrounding blocks, which represent the potential catchment area. The reason for the 2-mile radius is that the 2-mile is a 15-minute driving distance in the urban environment. The previous section has considered the catchment of walking-oriented customers by drawing a walking-distance buffer, and the driving-distance buffer at this micro-level evaluation further includes the driving-oriented customers for the new store.

Fig. 10
The map indicates the blocks within a 2-mile buffer around each target sites. The zoomed image indicates the target site block in.
The map also labels the locations of the potential sites.

Census block is a polygon feature, which is not very helpful when calculating distance in for applying the approach of micro-level evaluation, so data processing is required. I use the “Feature to Point” tool to find the centroid of each block polygon. The centroid contains all the attributes from the polygon, and we can use this new point layer to finish the later analysis. Census block is a polygon feature, which is not very helpful when calculating distance, so I use the “Feature to Point” tool to find the centroid of each block polygon. The centroid contains all the attributes from the polygon, and we can use this new point layer to finish the later analysis.

Fig. 11
Centroid contains the information of each census block.

Approach 1: Site potential interaction

My original approach is to use the interactive potential based on the interactivity index formula to find out the level of potential interaction between the block that contains target sites and the other surrounding blocks. The target sites with a higher cumulative interactivity index would be the most suitable location. The formula of the potential interactivity is presented below. Where the V is potential, P stands for population and d_{ij} is the Euclidian distance from the centroids of the surrounding blocks j to the centroid of the target site blocks i. The 0.01 is to simplify the final index and make the results more comfortable to read.

$$V_{target\ site} = 0.01 \times \sum_{j=1}^{n} \frac{P_i P_j}{d_{ij}}$$

However, the value of the interactive potential would be significantly pulled up when the target site block has a large population. In other words, if the target site block contains a large proportion of vacant space and a smaller population, the site would have a low interactive potential even though the surrounding blocks have a much larger population base. This circumstance may cause a problem for business site selection: if neighborhoods with large populations surround a vacant target site, this site should be valuable.

Approach 2: Surrounding population

To solve the problem from approach 1, I proposed this second approach: directly count the population of the surrounding blocks. The block with vacant spaces may be surrounded by high population density neighborhoods, which can also represent the great customer base for the new store.

The population count is a most straightforward method; other than this, weighted the population count at target site blocks and surrounding blocks can also relieve this problem.

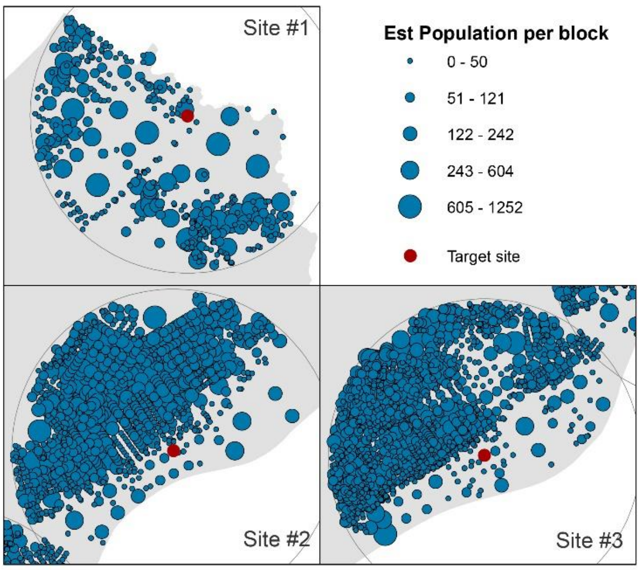


Fig. 12
The proportional map indicates population distribution around each target sites. The population around site #2 and site #3 are significantly larger.

Outcomes

The two approaches in the micro-level analysis provide different conclusions. Site #1 at Far Northeast Philadelphia and Site #2 at Kensington are the potential locations for the new grocery store at Philadelphia. Although Site #3 ranks in the meddle in both approaches, it just reflects that this site has an excellent balance between interactive potential and customer foundation. The table below includes the results of both methods. The evaluation also derived noticeable characteristics of each site and is presented in the table.

Target site	Approach 1	Approach 2	Site characteristics
Site #1	**16.4**	29530	Site is near to population concentrated districts like residential area. Trends to attract more walking-customers.
Site #2	6.09	**98993**	Site has vast vacant spaces for parking and construction. More attractive Driving-oriented customers, so this site can also consider having a gas station.
Site #3	10.97	95367	Site with a large interactive potential and a high population surrounding.

6. Conclusion

The overall goal of this project is to identify a suitable site for opening a new affordable grocery store in Philadelphia. The new grocery store location should contain both a large marketplace and available parking space. To achieve this aim, I use a top-down workflow to apply a multi-criteria analysis on the study area and eventually select a potential site based on different micro-level approaches. The outcomes also summarize the characteristics of each possible location, and the decision-maker can utilize the results to make the final decision of the new store location.

The project has several limitations that need to improve. Firstly, the project assumes the population is evenly distributed within continuously spatial units, which may reduce the accurateness of the analysis. For the next step developments, I would prefer to include more business-side data (sales volume, marketplace area, etc.) to conduct conjoint and segmentation analysis; therefore, the project will be more comprehensive and feasible.

Reference

Church, R. L., & Murray, A. T. (2009). *Business site selection, location analysis, and GIS* (pp. 20-22). Hoboken, NJ: John Wiley & Sons.

Hamrick, K. S., & Hopkins, D. (2012). The time cost of access to food–Distance to the grocery store as measured in minutes. *International Journal of Time Use Research*, *9*(1), 28-58.

Murad, A. A. (2003). Creating a GIS application for retail centers in Jeddah city. *International Journal of Applied Earth Observation and Geoinformation*, *4*(4), 329-338.

Rikalovic, A., Cosic, I., & Lazarevic, D. (2014). GIS based multi-criteria analysis for industrial site selection. *Procedia Engineering*, 69(12), 1054-1063.

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*, *40*, 191-198.

Van Eck, J. R., & De Jong, T. (1999). Accessibility analysis and spatial competition effects in the context of GIS-supported service location planning. *Computers, environment and urban systems*, *23*(2), 75-89.

Yang, Y., & Diez-Roux, A. V. (2012). Walking distance by trip purpose and population subgroups. *American journal of preventive medicine*, *43*(1), 11-19.