**Final Project Report**

Title: Using fuzzy suitability analysis to determine retail locations in the Metropolitan counties, MN

Notice: Dr. Bryan Runck

Author: Maisong

Date 4/29/2021

**Project Repository:**<https://github.com/msongfrancis/gis5572project.git>

**Abstract**

GIS has been used to solve location problems such as finding the best location for retail, emergency services or warehouses (Church 2002). In retail, demographics of existing and ideal costumers are also considered when determining the next retail location (Trubint et al. 2006). Costco, a warehouse style store, has nine locations in the seven metropolitan counties of Minnesota. This study aims to determine suitable locations for a new Costco location based on social demographics, accessibility, and proximity to competing business and existing Costco locations. The results show that the peripheral tracts have the highest suitability score when considering all the demographic and proximity factors. There are also a few tracts within the city that appear to have relatively high suitability score. This is suspected to be caused by high scores in a demographic parameter like income and accessibility to roads since these are located near downtown areas. This model can be improved by incorporating more detailed demographic information, where users are coming from, networks for time cost considerations, and resources available at each existing location.

**Problem Statement**

When determining where to build their next retail locations, businesses consider many factors. A few are accessibility, existing services, and demographics of their ideal and existing consumers (Trubint et al. 2006). This study aims to determine potential locations for a new retail Costco Store using suitability analysis.

*Table 1. Components needed to find suitable Costco locations.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **#** | **Requirement** | **Defined As** | **Spatial Data** | **Attribute Data** | **Dataset** | **Preparation** |
| 1 | Areas close to major roads like highways | Accessible roads – will be used to access proximity to major roads/highways | Lines |  | [Mn Geospatial Commons](https://gisdata.mn.gov/dataset/us-mn-state-metc-trans-fnctnl-cls-rds) |  |
| 2 | Areas available for business development | Regionally planned areas for different categories of land use like business development. | Raster | Planning Code | [Mn Geospatial Commons](https://gisdata.mn.gov/dataset/us-mn-state-metc-plan-pland-land-use) |  |
| 3 | Where the stores already area and where competing stores are located | Locations of other businesses that offer similar services like Target, Walmart, and Sam’s Club and existing Costco stores | Points | Lat, long  Business name | Google places API | Transform x,y to point features. |
| 4 | Social demographics about neighborhoods | Social information about each census tract |  | Age, number in household, family size, percentage in work force, median household income | IPUMS- ACS survey | Remove unnecessary columns, convert datatype, calculate rates. |

**Input Data**

The demographic data required are metrics regarding median household income, household size, labor status in the work force and age. The demographic data will be obtained on a census tract level and the census tract boundaries will be sourced from the U.S. Census Bureau. As for the competing locations and existing locations, the coordinates will be obtained from Google using their Google places API. The roads and business development zones will be sourced from MN Geospatial commons as shapefiles provided by the Metropolitan Council.

*Table 2. Data required to analyze problem statement.*

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Title** | **Purpose in Analysis** | **Link to Source** |
| 1 | MN Major Roads and Highways | Accessible roads – will be used to assess proximity to major roads/highways | [Mn Geospatial Commons](https://gisdata.mn.gov/dataset/us-mn-state-metc-trans-fnctnl-cls-rds) |
| 2 | Regionally planned areas | Regionally planned areas for different categories of land use like business development/commercial usage. | [Mn Geospatial Commons](https://gisdata.mn.gov/dataset/us-mn-state-metc-plan-pland-land-use) |
| 3 | Competing Business Locations | Locations of other businesses that offer similar services like Target, Walmart, and Sam’s Club | Google places API |
| 4 | Social demographics | Median HH income, AVG HH Size, AVG Family Size, Population Age 21+, percentage in labor force. | [IPUMS](https://www.census.gov/data.html) |
| 5 | Census Tract | Boundaries for mapping and used for site selection. May calculate centroid for network analysis for accessibility. 2018 | [US Census Bureau](https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html) |
|  | Existing Costco Locations | Examine demographics around existing Costco and how far they are from each other. | Google places API |

**Methods**

*Demographics*

First unnecessary demographic data was removed from the tables and reformatted to have clear headers. The tabular demographic data was then joined to the census tract shapefile. In ArcGIS Pro, the only rate calculated was population over 21 for each tract. The data already had rates like average household size, average median family income, average median non-family income, and average family size. Afterwards, each field was converted to a raster and the fuzzy membership tool was applied to rescale from 0-1 based on preference of values (Fig 1).

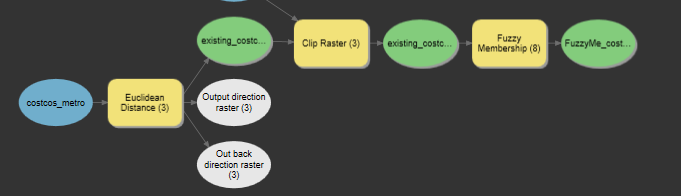
Figure 1. Workflow for transforming demographic data into layer for fuzzy suitability.



*Proximity Layers*

A raster layer was computed for proximity to major roads and highways, existing Costco warehouses, and competing stores by considering a 100m cell and computing Euclidean distance from each point. The fuzzy membership tool was applied to each raster with the membership with a Large membership type for Costco and competing because distance further from these features are preferred. For distance to roads, the small membership type was applied because being closer to major roads and highways were preferred for accessibility (Fig 2).

Figure 2. Workflow for proximity to features to create layer for fuzzy suitability.

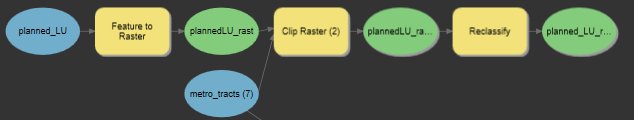




*Regionally planned areas*

To create a layer considering the regionally planned areas devoted to commercial and business, I first converted the features to a raster based on the land use classification code. I then reclassified the raster, so business areas had a value of 1 and the others had a value of 0. The raster was then clipped to the metro census tracts (Fig 3).

*Figure 3. Workflow to add planned land use to suitability analysis.*



*Calculating fuzzy suitability scores*

Once all the layers were produced and on the same scale (0-1), I used the fuzzy overlay tool to compute a fuzzy suitability layer. However, I also read that the fuzzy overlay tool would not be best for categorical data like the planned land-use layer (ESRI n.d). I chose to also do a weighted sum with the proximity to existing Costcos weighted as a 3 and demographics as 2. Everything else had a normal weight. (Fig 4).

*Figure 4. Fuzzy overlay and weighted sum tool workflow to produce fuzzy suitability score. A) fuzzy overlay. B) Weighted sum.*

|  |  |
| --- | --- |
| a) | b) |

**Results**

The fuzzy suitability score calculated using the fuzzy overlay tool ranged from 0-0.8342 (fig 5). In comparison the score calculated from the weighted sum ranged from 5.04-13.91. The majority of the fuzzy overlay results had a value of 0, but the areas that had higher scores showed it was areas near the edges (fig 5). Only areas that corresponded with the planned land use for business and commercial were given a score other than zero. When the land-use was not considered, the fuzzy overlay results showed more suitable areas near the periphery of the metropolitan counties area (fig 6). Overall the areas with the most suitable scores are located at the edges of the metropolitan counties area after consideration of all factors.

The weighted sum tool showed different results as well based on if the factors were equally weighted or weighted differently. In the unweighted, the same pattern of higher suitability scores near the periphery were visible with the lower suitability being smoothly dispersed from the center to the edges (fig 8). When the weights were considered, the scores were of course higher, but more areas had a higher suitability score (fig 97. Taking the averages of the calculated weighted sum scores for each tract, the higher suitability scores are actually present in the center of the Minneapolis and St. Paul area, as well as the outer tracts of the metro counties (fig 9).

*Figure 5. Fuzzy suitability scores calculated using the fuzzy overlay tool in ArcGIS Pro for potential Costco retail location suitability in the metropolitan counties of Minnesota.*

Diagram

Description automatically generated

*Figure 6. Fuzzy suitability scores calculated using the fuzzy overlay tool without Planned Landuse consideration in ArcGIS Pro for potential Costco retail location suitability in the metropolitan counties of Minnesota.*

Diagram

Description automatically generated

*Figure 7. Fuzzy suitability scores calculated using the weighted sum tool in ArcGIS Pro for potential Costco retail location suitability in the metropolitan counties of Minnesota with proximity to existing services and demographics weighted.*

Map

Description automatically generated

*Figure 8. Fuzzy suitability scores calculated using the weighted sum tool in ArcGIS Pro for potential Costco retail location suitability in the metropolitan counties of Minnesota with all factors equally weighted.*

Map

Description automatically generated

*Figure 9. Average fuzzy suitability scores for each census tract calculated using the weighted sum tool in ArcGIS Pro for potential Costco retail location suitability in the metropolitan counties of Minnesota*

Map

Description automatically generated

**Results Verification**

To verify the results, I looked at the metrics at existing Costco locations to make sure they met the criteria as well. Most of the metrics were met except for the household size for the St. Louis Costco location being 1.67 and the average family size being 2.61. However, the tracts surrounding this location did all meet the demographic criteria with a score of at 9 which is more than many of the other scores for other Costco locations.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **name** | **address** | **avg\_fam\_size\_2** | **avg\_hh\_size\_2** | **med\_inc\_fams\_2** | **per\_over\_21** | **med\_age\_2** | **Tract suitability score** |
| Costco Business Center | 3311 Broadway St NE, Minneapolis, MN 55413 | 2.89 | 2.8 | 46714 | 0.817126 | 24.2 | 8.775961 |
| Costco Wholesale | 5801 W 16th St, St Louis Park, MN 55416 | 2.61 | 1.67 | 111250 | 0.920823 | 32.8 | 8.307104 |
| Costco Wholesale | 1431 Beam Ave, Maplewood, MN 55109 | 3.05 | 2.5999999 | 114722 | 0.76533 | 44.1 | 8.624203 |
| Costco Wholesale | 995 Blue Gentian Rd, Eagan, MN 55121, | 3.29 | 2.75 | 97361 | 0.727923 | 31.8 | 8.897712 |
| Costco Wholesale | 11330 Fountains Dr, Maple Grove, MN 55369 | 2.99 | 2.28 | 118838 | 0.744785 | 36.5 | 8.786936 |
| Costco Wholesale | 12011 Technology Dr., Eden Prairie, MN 55344 | 3.18 | 2.5599999 | 77176 | 0.709548 | 31.6 | 9.392391 |
| Costco Wholesale | 7070 Tamarack Rd, Woodbury, MN 55125 | 2.95 | 2.3 | 92750 | 0.743585 | 34.6 | 8.769773 |
| Costco Wholesale | 12547 Riverdale Blvd, Coon Rapids, MN 55448 | 2.91 | 2.45 | 92872 | 0.742279 | 43.8 | 8.695862 |

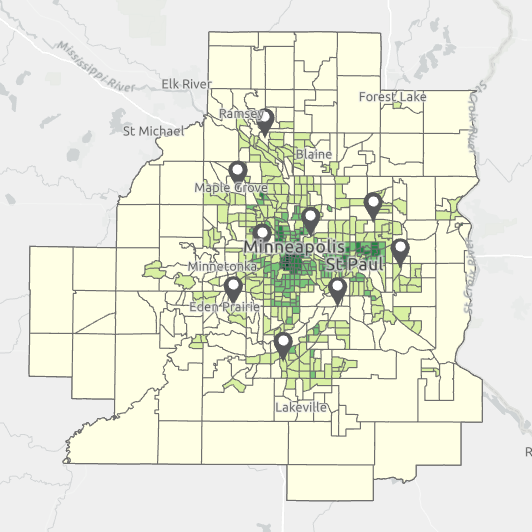
**Discussion and Conclusion**

Figure 8. Population density per square mile in the Metropolitan Counties of Minnesota.

Demographic metrics like average household size, average family size, median income, and population percentage over the age of 21, and proximity metrics considering location near existing and competing businesses and major roads were used in this fuzzy suitability analysis. The results showed the most suitable areas were in the peripheral metro. There were also some suitable census tracts within the city, however because they were already close to existing locations, their higher score could be attributed to proximity to the roads which downtown areas are typically very accessible. The income of folks who live downtown may also be higher, thus contributing to the higher score.

When the results are compared in correlation with the population density, areas that appear to have low population density were also areas with high fuzzy suitability scores (fig 8). Furthermore, the existing Costco locations are in the areas with higher population density already in the metropolitan counties. This model is limited by the factors and information. It does not consider other demographic variables that are available through the census. Proximity is also considered in a Euclidean nature rather than an accurate realistic measure with consideration of time costs. I also think there could more accurate information that businesses have can help improve this model. I was making general assumptions such as areas closer to roads are more preferrable and median income higher than 50,882 is preferred. However, businesses like Costco may have their own information about the users of their services which would produce more accurate results than these.

This model in this does show that the existing Costco areas were in the midpoint of the suitability range, suggesting that these demographic and proximity parameters can be useful in determining potential business locations. Observing the demographic metrics in tracts of existing Costco locations on a binary suitability, most locations would be considered suitable except for the Costco business center which serves to be target towards inventory for businesses.

One thing I struggled with most was the fuzzy overlay tool. It seemed that most areas resulted in zeros as the computed fuzzy suitability score. I did not understand how it was computing it when I was using the AND overlay type. I expected it to produce results like in the weighted sum where even if it had a low score, it would not necessarily be given a zero because it scores well in other parameters.

Future studies may improve this fuzzy suitability analysis by also considering the resources available at each Costco store and accessibility by road with the cost of time considered. I think understanding where folks are coming from and which stores, they are likely to go to base on their location would help businesses determine where they should put a new location.

**References**

Church, R.L. 2002. Geographic information systems and location science. *Computers and Operations Research* 29:541-562.

Costco Wholesale Corporation. (2017). Seeking Alpha. https://seekingalpha.com/symbol/COST/earnings/transcripts

Cortland, P. 2020. Are you part of the middle class? This calculator can tell you. Considerable. https://www.considerable.com/money/economy/what-is-middle-class-in-every-state/

ESRI (n.d). Applying fuzzy logic to overlay rasters. ArcGIS Pro. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/applying-fuzzy-logic-to-overlay-rasters.htm

Trubint, N., Ostojic, L., and Bojovic, N. 2006. Determining an optimal retail location by using GIS. *Yugoslav Journal of Operations Research*. 16.2:253-264.

**Self-score**

*Fill out this rubric for yourself and include it in your lab report. The same rubric will be used to generate a grade in proportion to the points assigned in the syllabus to the assignment.*

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Description** | **Points Possible** | **Score** |
| **Structural Elements** | All elements of a lab report are included **(2 points each)**:  Title, Notice: Dr. Bryan Runck, Author, Project Repository, Date, Abstract, Problem Statement, Input Data w/ tables, Methods w/ Data, Flow Diagrams, Results, Results Verification, Discussion and Conclusion, References in common format, Self-score | 28 | **28** |
| **Clarity of Content** | Each element above is executed at a professional level so that someone can understand the goal, data, methods, results, and their validity and implications in a 5 minute reading at a cursory-level, and in a 30 minute meeting at a deep level **(12 points)**. There is a clear connection from data to results to discussion and conclusion **(12 points)**. | 24 | **22** |
| **Reproducibility** | Results are completely reproducible by someone with basic GIS training. There is no ambiguity in data flow or rationale for data operations. Every step is documented and justified. | 28 | **25** |
| **Verification** | Results are correct in that they have been verified in comparison to some standard. The standard is clearly stated **(10 points)**, the method of comparison is clearly stated **(5 points)**, and the result of verification is clearly stated **(5 points)**. | 20 | **15** |
|  |  | 100 | **90** |