Geospatial Analysis for Flagship Store Site Selection in Metro Vancouver

Jessica Sun

Introduction

Opening a flagship store is a financial risk that requires much material, labor, and time investment. As such, when expanding, a company must carefully consider all aspects of the new store's location and surroundings to ensure that it is a profitable endeavor. The objective of this research project is to recommend five of the best candidates, and ultimately select one top choice, for a new outdoor gear retail and rental store in Metro Vancouver. This will allow the company to expand and open a new branch in a way that is both feasible and allows the company to make the most profit.

The approach to the analysis is to weigh four main categories: accessibility, customer base, foot traffic, and feasibility. The analysis is done based on 20 candidate sites, where the surroundings and exact location of each site is used to calculate relevant data to compare the sites. Once all data is collected, it is normalized, weighed by category, and scaled, and the top 5 of these scores correspond to the top 5 candidate store locations chosen.

Data

The datasets used are: Population, Median Household Income, Dissemination Areas, Points of Interest (POI), Transit Points, Streets, Land Use, Land Cover, and Elevation. All datasets are projected to NAD 1983 UTM Zone 10N for analysis. The following table describes the basic characteristics of each dataset, and its reference.

Table 1: *Characteristics and References for each Dataset*

Dataset	Reference	Year	Spatial Data Type	Source
Population	7	2022	Vector	Government-collected
Median Household Income	7	2022	Vector	Government-collected
Dissemination Areas	6	2021	Vector	Government-collected
Points of Interest	3	n.d.	Vector	Crowdsourced
Transit Points	3	n.d.	Vector	Crowdsourced
Streets	3	n.d.	Vector	Crowdsourced
Land Use	4	2016	Vector	Government-collected
Land Cover	5	2020	Raster	Government-collected

Digital Elevation Model	8	2000	Raster	Government-collected

Methods

My first step was to look at all my datasets to figure out what kind of analysis I could do on each one to gain valuable insight about the location of each candidate store. I decided the analyses suitable for each dataset and each candidate as follows: 1 kilometer and 5 kilometer buffer intersection with Dissemination Areas for Population, 5 kilometer buffer intersection with Dissemination Areas for Median Household Income, 2 kilometer buffer for count of Points of Interest, Euclidean distance to nearest bus stop (Transit Points), 2 kilometer service area (Streets), Land Use at the precise location, 5 kilometer buffer for percentage Paved Land Cover, and 5 kilometer buffer for average slope (Elevation).

I then grouped these analyses into categorical criterion, to make my final results easier to "weigh" against each other. The first category is "Accessibility", encompassing nearest bus stop and size of service area. The second is "Customer Base", encompassing population and household income. The third is "Foot Traffic", encompassing points of interest and average slope. Lastly, the feasibility category encompasses the percentage of paved land. I decided to not weigh my land use analysis and instead automatically eliminate all candidates which didn't fall under "Retail and Other Commercial (S200)", because it appears to be the only category that allows for retail operations such as this flagship store.

Before starting with any analysis, I made sure my data was uniform by projecting all datasets into NAD 1983 UTM Zone 10N Projection. This projection is ideal for this project because it puts all analysis in the perspective of projected meters, and also focuses on western North America, which is where Metro Vancouver is situated. I also added a spatial context to the CSV files by exporting the points of interest as a point feature class, and creating a join by attribute for the population and median household income datasets to the dissemination area dataset by DAUID.

The first set of analyses I did was for the "Accessibility" category. This category focuses on how easy it is to reach each potential location. Because the primary form of transportation in Metro Vancouver is by bus, an accurate measure of accessibility is to look at how close each location is to a bus stop. Additionally, service area analysis reveals information about accessibility via the general road network. Smaller distance to bus stops and larger service areas are considered "good" metrics. To find the distance to the nearest bus stop from each candidate location, I selected points labeled as "bus stops" (filtering out railway stations, ferry terminals, etc.) from the Transit Points dataset, and created a new layer with only those points. I then found the closest stop to each location in meters using the planar method, and recorded the distance for each candidate. To find the service area within 2 kilometers of each candidate, I first created a new feature dataset and imported the Streets dataset into it. I then created a network dataset from the newly created feature dataset (with no elevation model), and built a network. To actually find the service areas, I created a service area analysis layer and added the candidates as facilities, then solved the analysis to generate polygons representing the accessible areas within 2 kilometers along the road network for each candidate site. I found the area (in square meters) for each polygon by visiting the attribute table for this newly generated layer, and recorded the values.

Next, I completed the analyses for the “Customer Base” category, which focuses on the potential customers for the store. Both the number of customers (population) and their income are key measures, with larger populations and higher incomes considered “good” metrics. To find the population within 1 kilometer and 5 kilometers of each candidate location, I created pairwise buffers at both distances and spatially joined them with the Dissemination Area layer using the “intersect” property. I configured the join to sum the population fields for each buffer and recorded the results. To find the median household income, I used the same 5 kilometer buffer and spatially joined it with the Dissemination Area layer, this time calculating the median of the income field for each candidate.

The next category of analysis was “Foot Traffic,” which focuses on how often people might be nearby, allowing new customers to discover the store and ensuring proximity to other frequently visited locations. I measured this with points of interest and average slope: a higher number of nearby points of interest suggests more potential customers, while higher average slope indicates more mountainous terrain, attracting the outdoor gear market. To find the number of points of interest within 2 kilometers of each candidate, I created a pairwise buffer of 2 kilometers and spatially joined it with the Points of Interest layer using the “contains” property, recording the resulting point count. To calculate average slope within 5 kilometers, I created a slope raster from the elevation dataset (with a z-factor of 1 and output in percent rise), then extracted the mean slope value for each buffer using zonal statistics as a table, and recorded the results.

For my “Feasibility” analysis, I focused on land cover, as building a store at a location that is already paved would save expenses and time. To find the percentage paved land cover within 5 kilometers of each candidate location, I first created a new raster layer from the Land Cover layer, using the map algebra expression “"MetroVan_LandCover.tif" == 2” (the number 2 corresponds to the paved land cover type in the dataset), so that this new layer would have paved land cover labeled 1 (true), and every other land cover type labeled 0 (false). I then used this layer along with my pre-existing 5 kilometer buffer around the candidate locations, and found zonal statistics as a table, using the statistics type “sum” to calculate the number of paved pixels within each buffer zone. This works because paved land cover type is now associated with the value 1, so summing the values results in the sum of the number of paved pixels per zone. Lastly, I divided this paved pixel count by the total pixel count for each buffer area, and recorded this data.

The last part of my analysis was land use. This is important because it identifies what candidate locations are feasible (identified as "Retail"). To find the land use for each candidate, I sorted the Land Use dataset by fclass. I then did a spatial join of the Land Use layer with the candidates layer with the "within" operation to find the land use type at each location, and recorded this data. All locations ended up having either "Retail" or "Industrial" land use types. Upon checking the source data from Metro Vancouver Open Data Portal, I found that the "Industrial" category had the description of "...processing, manufacturing, warehousing and wholesaling activities." (4) This doesn't fit the description of the type of retail flagship store under analysis, so I eliminated all the "Industrial" land use category locations and kept only the "Retail" ones.

After recording all data, I decided on a weighting scheme for my categories. Accessibility and customer base were the two categories that stood out to me as most important because they determined the financial outcome and overall future of a potential new store. I weighed "Accessibility" and "Customer Base" each at 30% (Nearest Bus Stop (15%), Service Area (15%), Population within 1 kilometer (10%), Population within 5 kilometers (5%), Household Income (15%)). I weighed "Foot Traffic" at 20%, as a fairly important category in terms of new customers (Points of Interest (15%), Average Slope (5%)). Lastly, I weighed "Feasibility" at 10% (Percentage Land Cover Paved (10%)) as the least important category, because it is always an option to pave additional areas at an extra expense if all other categories score high. I then normalized data for each subcategory using a normalization method (1) such that Normalized Value = (Value-Min)/(Max-Min) for benefit criteria and inverted formula Normalized Value = (Max-Value)/(Max-Min) for cost criteria. Lastly, I multiplied all values by their respective subcategory weight, and added up all final values for each candidate.

Results

Based on the normalized weighting method described in the previous section, I found that the candidate locations with the largest final values were candidates 18, 20, 16, 15, and 8, in decreasing order. Their scores were 71.72, 54.30, 53.92, 52.41, and 47.72, respectively. The following table (Table 2) outlines the comparisons between the population data, distance to nearest bus stop, average percentage slope, percentage paved land cover, number of points of interest, and land use type for the top five candidates.

Table 2: *Comparison of Key Suitability Criteria for Top 5 Candidate Store Locations*

Candidate No.	**18**	**20**	**16**	**15**	**8**
Population of intersecting DAs within 1km	24686	22182	28980	20453	13304
Population of	191278	210452	196717	160278	199769

intersecting DAs within 5km					
Distance to nearest bus stop (m)	32.5	41.34	126.73	133.81	128.91
Average % Slope within 5km	15.10	14.10	13.39	18.02	14.60
Percentage “Paved” within 5km	23.84	25.41	26.97	19.51	24.60
Number of POI within 2km	412	372	352	202	276
Land Use Type	Retail	Retail	Retail	Retail	Retail

The map displayed below (Figure 1) is a small-scale map of the top five candidate locations, overlaid with the Streets dataset and the Elevation dataset. It also displays points of interest within 2 kilometers of the candidates, and median household income (color-coded by natural break ranges of income) for each dissemination area intersecting within 5 kilometers of one of the top five candidates.

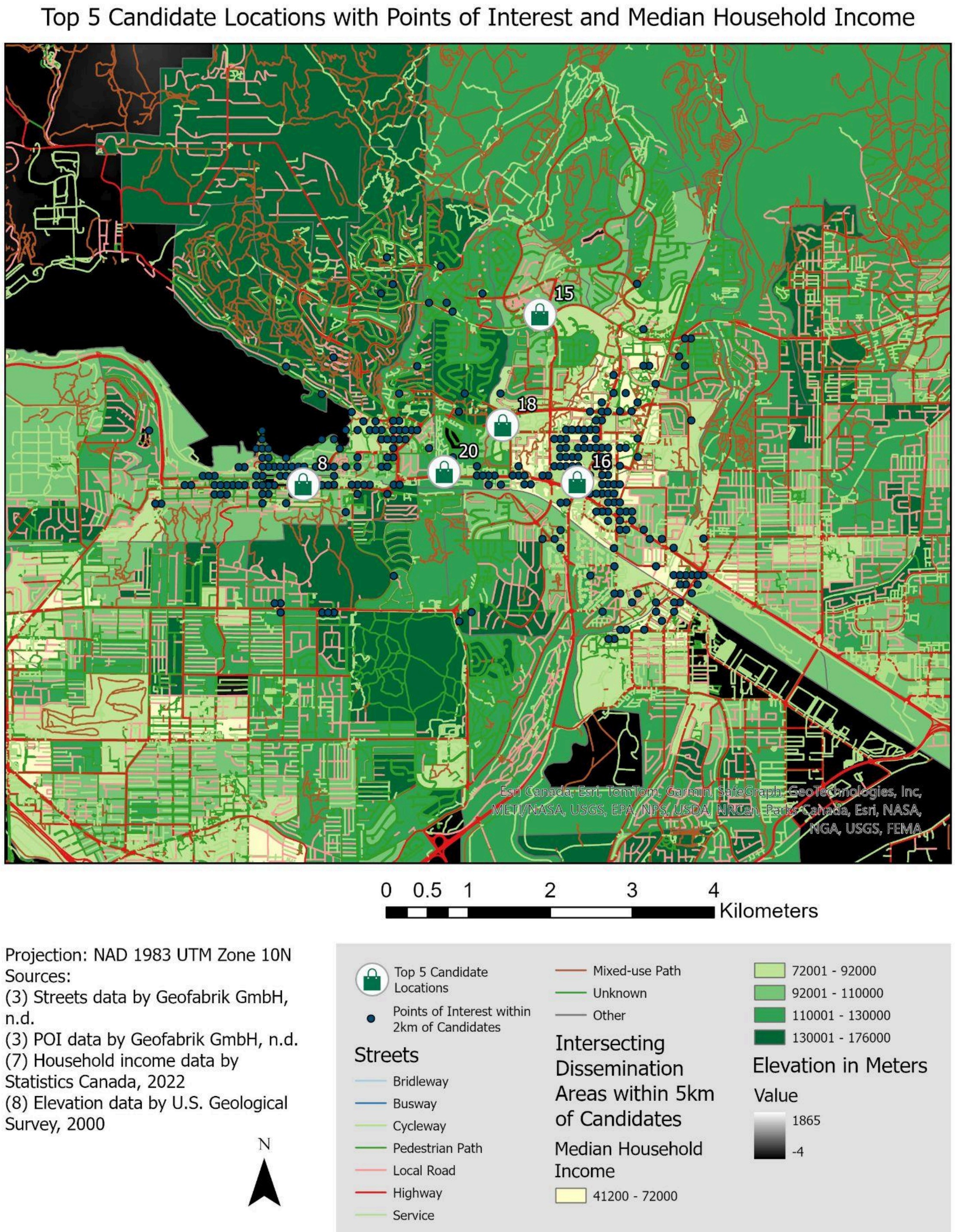

Figure 1: *Top 5 Candidate Locations with Points of Interest and Median Household Income*

The following map (Figure 2) is a large-scale map that zooms in on the top candidate (candidate 18). It includes the land use layer, all bus stops within a 2 kilometer buffer (Euclidean distance), and a polygon of 2 kilometers along the street network representing the service area analysis layer.

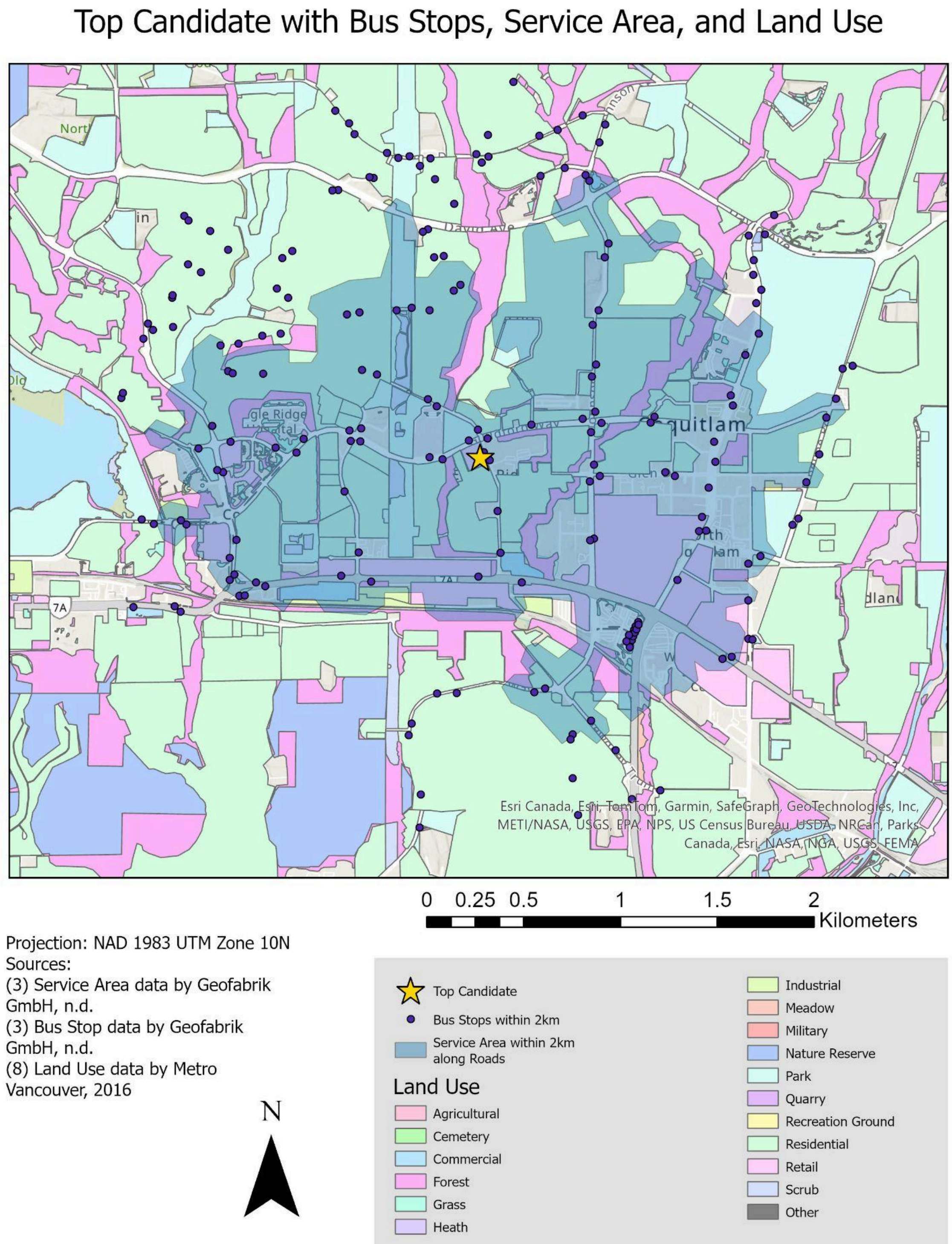

Figure 2: *Top Candidate with Bus Stop, Service Area, and Land Use*

Discussion

My conclusion that site 18 is the best candidate location is based on its high performance across nearly all major categories considered in this analysis. Candidate 18 had the shortest distance to a bus stop (32.5 meters), ensuring high accessibility for customers using the bus, which is the most common form of transportation in Metro Vancouver. It also had the highest number of points of interest within 2 kilometers (412 POIs), indicating potential for high foot traffic and incidental customer discovery. Although its percentage of paved land (23.84%) was slightly lower than some other top candidates, it was still substantial enough to suggest a feasible construction environment without requiring major additional infrastructure. While other candidates were competitive, site 18 demonstrated a balanced strength across all factors.

This study is relevant because it identifies, for this national outdoor gear retailer, an ideal store location in Metro Vancouver. It's a highly important study for the company, because a good location is a hugely influential factor in terms of the future income and profitability of a new flagship store.

There are, however, some weaknesses in my approach. One limitation is that my analysis relies purely on secondary geospatial data without site-specific field verification. This introduces potential errors if data is outdated, incomplete, or misclassified. For example, the Points of Interest dataset may not reflect recent business openings or closures. Similarly, assuming that proximity to bus stops automatically translates to higher customer volume may overlook real-world factors such as pedestrian infrastructure or bus frequency. Additionally, my judgment of the weighting scheme is subjective. Different weightings could lead to different top candidates. Furthermore, field visits to the top sites would allow for validation of accessibility, visibility, parking availability, and surrounding amenities—factors that are difficult to fully capture through geospatial analysis alone.

Conclusions

This project successfully identified the top candidate locations for a new flagship outdoor gear store through a systematic geospatial analysis approach. By carefully combining population, accessibility, foot traffic, and feasibility data and normalizing the results with a weighted scheme, candidate 18 emerged as the strongest recommendation, proving to have the best balance of important characteristics of the candidate locations.

Overall, geospatial analysis proved to be an effective first step for narrowing down candidate sites across a large metropolitan area. It allowed for objective, replicable comparisons across many variables that would be difficult to assess manually. However, the analysis could be greatly strengthened by a second phase involving field surveys or on-site inspections to validate the final shortlist. In future projects, I would consider supplementing GIS-based selection with additional qualitative criteria, such as visibility from major roads or adjacent business types, which are harder to capture quantitatively. Nevertheless, this project demonstrates that spatial analysis is a powerful tool for informed, data-driven site selection.

References

1 Bobbitt, Z. (2020, February 27). *How to normalize data between 0 and 100*. Statology. https://www.statology.org/normalize-data-between-0-and-100/

2 Esri. (2020). ArcGIS Pro (Version 2.6) [Computer software]. Environmental Systems Research Institute. https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview

3 Geofabrik GmbH. (n.d.). British Columbia - OpenStreetMap data extracts. Retrieved April 19, 2025, from https://download.geofabrik.de/north-america/canada/british-columbia.html

4 Metro Vancouver. (2016). *Land Use 2016 – Code Description.* Metro Vancouver Open Data Portal. Retrieved April 19, 2025, from https://open-data-portal-metrovancouver.hub.arcgis.com/datasets/metrovancouver::landuse-2016-code-description/about

5 Metro Vancouver. (2020). *Land Cover Classification 2020 (raster).* Metro Vancouver Open Data Portal. Retrieved April 19, 2025, from https://open-data-portal-metrovancouver.hub.arcgis.com/datasets/5dd153684b9b41249c0dcf09e79c9b25/about

6 Statistics Canada. (2021). Standard Geographical Classification (SGC) – Boundary files, 2021 Census. https://www12.statcan.gc.ca/census-recensement/2021/geo/sip-pis/boundary-limites/index2021-eng.cfm?year=21

7 Statistics Canada. (2022). Census Profile, 2021 Census of Population: British Columbia (Catalogue No. 98-401-X2021006). https://www150.statcan.gc.ca/n1/en/catalogue/98-401-X2021006

8 U.S. Geological Survey. (2000). USGS EROS Archive – Digital Elevation – Shuttle Radar Topography Mission (SRTM). Retrieved April 19, 2025, from https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm#overview