SA8912 Spatial Technologies in Strategic Planning

Assignment 1: Market Screening and Site Evaluation

Part 1 Market screening using Cluster Analysis (25 marks)

As of 2020, Dollarama has 156 stores in the Toronto CMA. The company is expanding and plans to add another 5 stores in 2024. In which areas in the CMA would you suggest Dollarama to consider setting up the new stores and search for suitable real estates?

Tasks:

1. Write a literature review on Cluster Analysis (3-4 pages, double spaced)

2. Perform a K-means Cluster Analysis in SPSS on the provided census data to identify the spatial clusters where "high" demand for Dollarama merchandise likely exists.

   Perform the cluster analysis twice: first, using the original variables; second, using weighted and standardized variables.

   - Use the following suggested census variables as classification variables (note: one of them need to be calculated by yourself from the original variables in the data set – known as "derived variable")

     (Feel free to choose a different set of classification variables, but do justify your choice.)

Table 1 Suggested variables

Variable label	Variable description	Assumption
im_11_16	Recent immigrants who arrived between 2011 and 2016	Tend to have low income
unemployed	Number of unemployed persons	Low/reduced income
renter	Home renter	Usually with low income
Avg_val	Average value of dwellings	Usually households with low income living in dwellings of low value
subsid	% of tenant households in subsidized housing	With low income
med_hh_inc	Median household total income $	
Hh_under50K	Number of private households with household income below $50K (this is a derived variable to be calculated by yourself).	Low income household

Experiment with K=4, k=5, and K=6 to find the most suitable number of clusters.

3. Join the SPSS data table (containing the cluster ID) with the MapInfo table (or the ArcGIS Table), and map the clusters.

4. Add the existing Dollarama stores to the map

5. Suggest areas for the 5 planned stores. (You may suggest opening fewer than 5 new stores if you find the CMA market is already saturated.)

6. Explain (1) your location suggestions and (2) the differences between the weighted-standardized cluster analysis and the un-weighted and un-standardized cluster analysis. (2-3 pages, excluding tables and figures)

Toronto CMA 2016 Census Variable Definitions

Database Field	Original Field
ctuid	Census tract ID
total_pop	Total population
Can_cz	Canadian citizens
im	Immigrants
Im_11_16	Recent immigrants who arrived between 2011 and 2016
Tot_vm	Total visible minority population
VM_SA	South Asian
VM_CHI	Chinese
VM_BLK	Black
VM_PILI	Filipino
VM_LA	Latin American
VM_arab	Arab
VM_SEA	Southeast Asian
VM_WA	West Asian
VM_KOR	Korean
VM_JAP	Japanese
unempl	Number of unemployed persons
Unemp_rate	Unemployment rate
owner	Home owner
renter	Home renter
Avg_val	Average value of dwellings
subsid	% of tenant households in subsidized housing
med_inc	Median personal income $
avg_inc	Average personal income $
med_hh_inc	Median household total income $
avg_hh_inc	Average household total income $
Total_HH	Total number of private household
hti_und5k	Under $5,000
hti5_10k	$5,000 to $9,999
hti10_15k	$10,000 to $14,999
hti15_20k	$15,000 to $19,999

hti20_30k	$20,000 to $29,999
hti30_40k	$30,000 to $39,999
hti40_50k	$40,000 to $49,999
hti50_60k	$50,000 to $59,999
hti60_80k	$60,000 to $79,999
hti80_100k	$80,000 to $99,999
hti100_125	$100,000 to $124,999
hti125_150	$125,000 to $149,999
hti150kov	$150,000 and over
pop_prv	Prevalence of low income household in 2015 based on after-tax low-income measure (%)
pop_25_64	Total population aged 25 to 64 years
Post_bach	population aged 25 to 64 years with University certificate, diploma or degree above bachelor level
bach_DEG	population aged 25 to 64 years with Bachelor's degree

Part 2 Construct a multivariate regression model for retail site evaluation (25 marks)

A Canadian retail chain operates 48 stores in the six major CMAs of Toronto, Vancouver, Montreal, Ottawa-Gatineau, Edmonton and Calgary. The dataset "MajorCMA_Store.sav" contains the following variables that represent the store attributes and trade area characteristics of the 48 stores:

- Store ID
- Province
- CMA
- Annual sales (average of 2018 and 2019 sales)
- Store size
- Number of direct competitors in trade area
- Number of secondary competitors in trade area
- Total number of households
- Population of 25-45 years of age
- Average household income
- Number of home owners
- Number of persons working at home
- Number of persons who moved in the past 5 year
- Number of new homes constructed between 2011 and 2016
- Number of recent immigrants

1. Write a literature review of multiple Regression Analysis (3-4 pages double spaced)
2. Using Annual sales as the dependent variable, and the 11 variables that are highlighted in green as the independent variables, to construct a regression model, using the "Enter" method. (Note: multiply "number of secondary competitors" by 0.6, to reduce its contribution to sales.)
3. Transform the variables into log form, and repeat Task 2 above.
4. Evaluate the two models by examining and interpreting the following:
    a. F-score and the associated significance level
    b. The R^2 of the model
    c. Significance levels for the coefficients of the independent variables
    d. Calculate and compare the projection errors (e.g., the differences between the model projected sales and the actual sales), including the average, minimum, maximum, range and standard deviation.

Template for evaluation and comparison of the two models

Model	F-score (with sig.)	R^2	Average error	Minimum error	Maximum error	Range	Standard deviation
Model 1							
Model 2							