3/22/2020

Data Science Professional Certificate Final Capstone Project

Analysis and Modelling of Loblaw Banners Selection for Private Investors

Yiteng Liang
COURSERA DATA SCIENCE COURSE BY IBM

1. Introduction/Business Problem

A private investor expects to open a grocery store or supermarket through joining the Loblaw franchises in Ontario, Canada. Loblaw is the top grocery store and supermarket franchising company in Canada that offers various franchise banners to be used for the independently owned franchise stores across Canada. Details of Loblaw franchise business opportunities can be found in this link.

Available franchise banners from Loblaw that are available for selection include

- No Frills™
- Independent City Market™
- Fortinos™
- Provigo™
- Valu-Mart™
- Your Independent Grocer™

These banners are different from each other and have their own unique features. Simply put, they are tailored for different types of communities or neighborhoods.

The private investor wants to open a store at a location where one of the banners from the list could be the best fit. Based on the private investor's description and preference, some other specific criteria should be considered when going through the analysis.

1. The best choice of location should be in the regions in western Ontario other than Toronto including Kitchener-Waterloo-Cambridge (KWC), London, and Windsor.
2. The pick of location and banner should be optimized to avoid harsh competitions within the same area.

Through this analysis, multiple appropriate locations with optimized selections of banners should be provided to the private investor to support his decision making. From the result, he should get a clearer and detailed insight of the locations and neighborhood features around them so that he can make the best decision about the business strategy.

2. Data

2.1 Data Understanding

The first task of this project is to figure out what differentiate the banners from each other. One of the straightforward strategies is to analyze the products and brands that each banner provides and find out what types of customer targeted by them. In order to do that, scraping the product and brand list for each banner and spending time on understanding the product-customer relationship is unavoidable. This could be extraordinarily time-consuming, and the return on the investment of time may not be satisfying eventually. However, it can be done from another point of view. Since Loblaw's business has been healthy for many years and its strategy of franchise expanding has been tested working by the real market, we have the confidence to believe that the existing locations of those stores have gone through meticulous determination after tremendous analysis performed by either Loblaw's own business teams or other consulting firms. Therefore, by analyzing the location data collected for the neighborhoods around the existing store locations, we are able to tell the uniqueness of each banner based on the similarities or differences of its neighborhood comparing to others.

The second task is to analyze multiple regions or cities and try to classify their neighborhoods in such a way that the banners are classified. By doing this, we can propose the top and second choices of banners for each of the neighborhoods in those regions of interest.

The last task, of course, is to help the private investor avoid potential risk of opening a store with duplicated function as one of the existing stores nearby. Competition with similar banners from other franchise companies than Loblaw is allowable.

2.1.1 Store/Banner Address Data

Running for the first task, we need to have a list of locations for the 6 banners. Locations of stores can usually be found by using the "store locators" provided by each store's website. Taking No Frills™ store locator for example, by inputing the region as Toronto, it shows all the locations with the corresponding addresses and working hours in the list. However, the temporary searched results, as attempted, are not easy to be scraped by Beautifulsoup pakage in the traditional way. Another approach is found and used here, that is, we can use a website scraping software called Parsehub. It remembers a sequence of selections and clicks needed for searching on the website before scraping the data outputed from the searches. The tutorial can be referred here.

From reading some basic descriptions of the banners, it is found that Provigo™ is only used for retailers based in Quebec. Since Quebec does not fall into the private investor's region of interest, Provigo™ is disregarded from our list. After using Parsehub to scrape the locations data of the rest 5 banners from their store locator websites, we obtained 5 csv files containing the information we need for geographical data analysis. Table 1 shows the partial data of the NoFrills address csv file.

Table 1: NoFrills Scraped Address Data Partial Example

Store Name	Address	Hours	In-store Only
ROCCO'S NOFRILLS TORONTO	200 Front St E, Toronto, Ontario M5A 4T9	Closed Now8:00 AM - 9:00 PM	In-Store Shopping Only
MATT'S NOFRILLS TORONTO	449 Parliament St, Toronto, Ontario M5A 3A3	Closed Now9:00 AM - 9:00 PM	In-Store Shopping Only

JOE'S NOFRILLS TORONTO	345 Bloor St E Unit 1A, Toronto, Ontario M4W 3J6	Closed Now8:00 AM - 9:30 PM	In-Store Shopping Only
DAVE & CHARLOTTE'S NOFRILLS TORONTO - GAS BAR	449 Carlaw Avenue, Toronto, Ontario M4K 3H9	Closed Now8:00 AM - 10:00 PM	In-Store Shopping Only
DAVE & CHARLOTTE'S NOFRILLS TORONTO	449 Carlaw Ave, Toronto, Ontario M4K 3H9	Closed Now8:00 AM - 10:00 PM	nan
JOE'S NOFRILLS TORONTO	900 Dufferin St, Toronto, Ontario M6H 4A9	Closed Now8:00 AM - 10:00 PM	nan

These address data sets should be cleaned and reorganized for our further usage. First, "hours" and "in-store only" features are not very important in our location analysis, so they will be removed from the tables. Then, the column names need to be formatted as "Name" and "Address". Second, the "Address" strings may need to be reworded or separated into different columns if necessary. In the data preparation section, coordinates need to be found for these addresses which are going to be passed to **Foursquare** to retrieve location data for further exploratory analysis.

2.1.2 Neighborhood Address Data

Either for analyzing the neighborhoods of the store locations found for the banners or the neighborhoods in the private investor's regions of interest (ROI), we need the geographical data around those points of interest. In this analysis, we will only use location data retrieved using **Foursquare** API. The location data from Foursquare only includes information like restaurants, coffee shops, fitness centers, airports, and so on. This location data is good enough to help figure out what type of life in a neighborhood. As a brief analysis, we will not go very deep into analyzing ages, income level, education, population or other demographics.

Toronto neighborhoods analysis as well as location data can be carried over from this report (Part 1 - Data and Part 2 - Analysis), in which the neighborhoods in Toronto are classified into clusters based on the Foursquare location data. We can use the similar method to analyze neighborhoods in Kitchener-Waterloo-Cambridge (KWC), London, and Windsor areas. The postal codes within these areas can also be found in this Wikipedia link. However, unlike Toronto, all the postal codes in the western Ontario start with 'N', so we need to extract only the postal codes for the three regions above from the site. Beautifulsoap package for data scraping will be used here to handle this task. Table 2 shows the first five records of the Toronto location data carried over from the reports, and Table 3 shows those for the ROI. Since the Toronto address data is carried over, it has the corresponding coordinate found for each postal code. In the next section, the ROI address data will be cleaned, and its coordinates will be found for all postal codes.

Table 2: Toronto Address Data First Five Records

Postcode	Borough	Neighborhood	Latitude	Longitude
M1B	Scarborough	Rouge, Malvern	43.80668	-79.19435
M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.78453	-79.16049
M1E	Scarborough	Guildwood, Morningside, West Hill	43.76357	-79.18871
M1G	Scarborough	Woburn	43.77099	-79.21691
M1H	Scarborough	Cedarbrae	43.77313	-79.23947

Table 3: ROI Address Data

Postcode	Region
N1A	Dunnville
N2A	Kitchener East
N3A	New Hamburg (Baden)
N4A	Not assigned
N5A	Stratford North

2.2 Data Preparation - Clean and Find Coordinates

For the scraped data, we need to clean the data and also find the coordinates for the store/banner addresses and neighborhoods in our regions of interest (ROI).

2.2.1 Getting Coordinates for Store/Banner Addresses

First, let's clean the stores dataset based on the requirements stated at the end of Section 2.1.1, and find the coordinates for the stores/banners. A tag column is added to indicate which banner the store uses. Table 4 shows part of this dataset.

Table 4: Banner Coordinates Data

Store Name	Address	Postcode	Latitude	Longitude	Banner
MATT'S NOFRILLS TORONTO	449 Parliament St, Toronto, Ontario	M5A 3A3	43.6555	-79.3626	nofrills
DAVE & CHARLOTTE'S NOFRILLS TORONTO	449 Carlaw Ave, Toronto, Ontario	M4K 3H9	43.6803	-79.3538	nofrills
...	...	...	...	...	...
TUFTS' VALU-MART	3259 Bayview Ave, Toronto, Ontario	M2K 1G4	43.7797	-79.3813	valumart
PETER'S VALU-MART MISSISSAUGA	1125 Bloor St E, Mississauga, Ontario	L4Y 2N6	43.6028	-79.5929	valumart
...	...	...	...	...	...

ALLAN'S YOUR INDEPENDENT GROCER	1900 Dixie Rd, Pickering, Ontario	L1V 1V4	43.8605	-79.1618	yourindgrocer
VOS' YOUR INDEPENDENT GROCER PORT PERRY	1893 Scugog Street, Port Perry, Ontario	L9L 1H9	44.1068	-78.9444	yourindgrocer

2.2.2 Getting Coordinates for Neighborhoods in Our Regions of Interest (ROI)

It is noticed that out of 180 records in Table 3, 74 have unassigned region. They will be ignored from our analysis.

Now, let's clean and reorganize the table so that it has a "Borough" column, in which the values will show whether it is Kitchener-Waterloo-Cambridge (KWC), London, or Windsor. At the mean time, the other boroughs not recognized as the ones in our regions of interest will be noted as 'Nan'. We will remove the "Borough" records with 'Nan' so that regions not in our interest will be discarded. Eventually, the regions we are going to pick should include the following information:

- Kitchener as "KWC"
- Waterloo as "KWC"
- Cambridge as "KWC"
- London
- Windsor

Table 5 shows part of the ROI coordinates dataset including our own defined Boroughs and Neighborhoods.

Table 5: ROI Coordinates Data

Postcode	Borough	Neighborhood	Latitude	Longitude
N2A	KWC	Kitchener East	43.4413	-80.4246
N2V	KWC	Waterloo Northwest	43.4764	-80.5842
N2T	KWC	Waterloo Southwest	43.453	-80.5692
…	…	…	…	…
N6A	London	London North (UWO)	42.9976	-81.2563
N6B	London	London Central	42.9835	-81.2386
…	…	…	…	…
N9K	Windsor	Windsor	42.2978	-82.8701
N8P	Windsor	Windsor (East Riverside)	42.3276	-82.9104

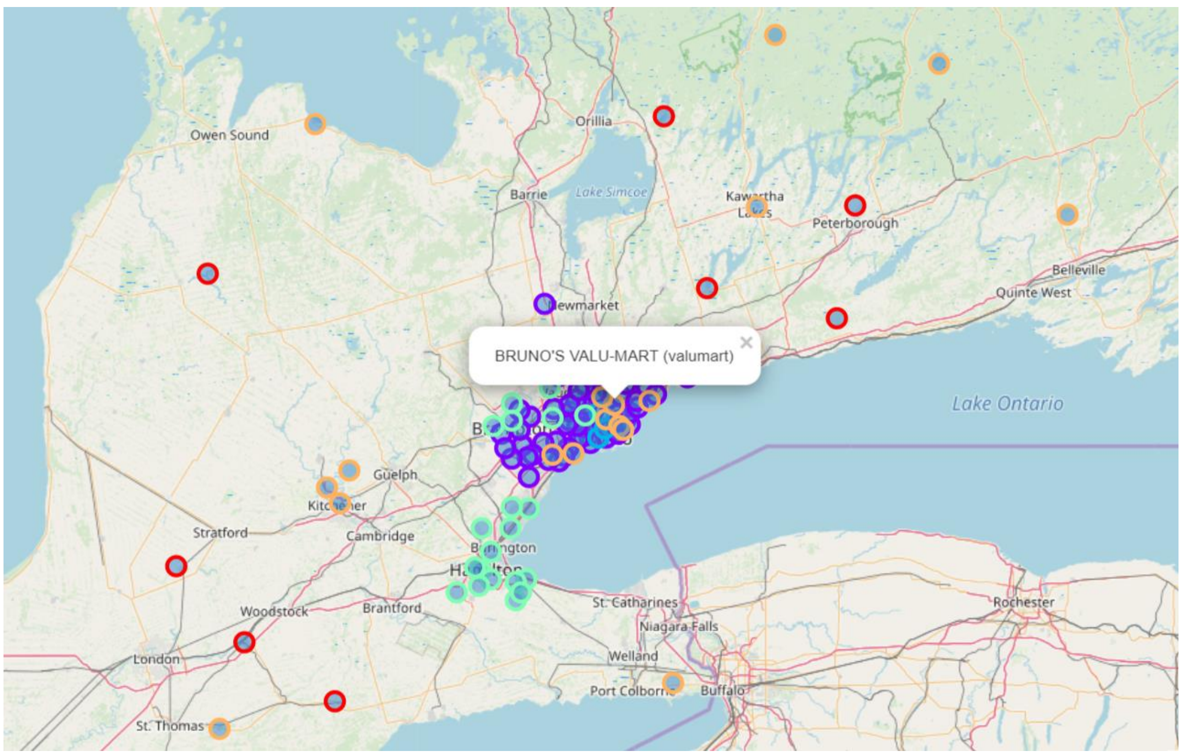
2.3 Data Preparation - Location Data through FourSquare

In this section, FourSquare API is used to retrieve location data for the stores and ROI's. Venues are found through FourSquare within a specified radius around the points of interest, which are inputs to the API as coordinates that were found for the stores and

ROI’s in the previous section. FourSquare will also return venue categories with the venues, and these categories will be used to study differences among those locations. From understanding these differences, we can separate these locations into groups and define proper features to describe each group uniquely.

2.3.1 Store (Banner) Location Data

First, the coordinates of the stores are plotted onto a map using Folium package to have an overlook of the banner geographical distributions. Figure 2.1 shows the geographical distributions of the banners. From the figure, it is observed that “NoFrills” stores are mainly gathered in Toronto with a high density, and “Your Independent Grocer” stores are the most dispersed in comparatively smaller cities. “Fortinos” stores are distributed along the areas near the lakeside from Hamilton to Toronto. “Value-Mart” stores have its businesses in both Toronto, Kitchener, as well as other dispersed locations.



(**Figure 2.1**: Geographical distributions of the banners)

After using the FourSquare API with a searching radius of 500m, nearby venues with categories are found for each store. Table 6 shows the top and bottom five number of venues found for the stores, and Table 7 shows the top 10 venue categories out of over 200.

Table 6: Store Venues Count

Neighborhood	Number of venues
CARMEN'S INDEPENDENT CITY MARKET	83

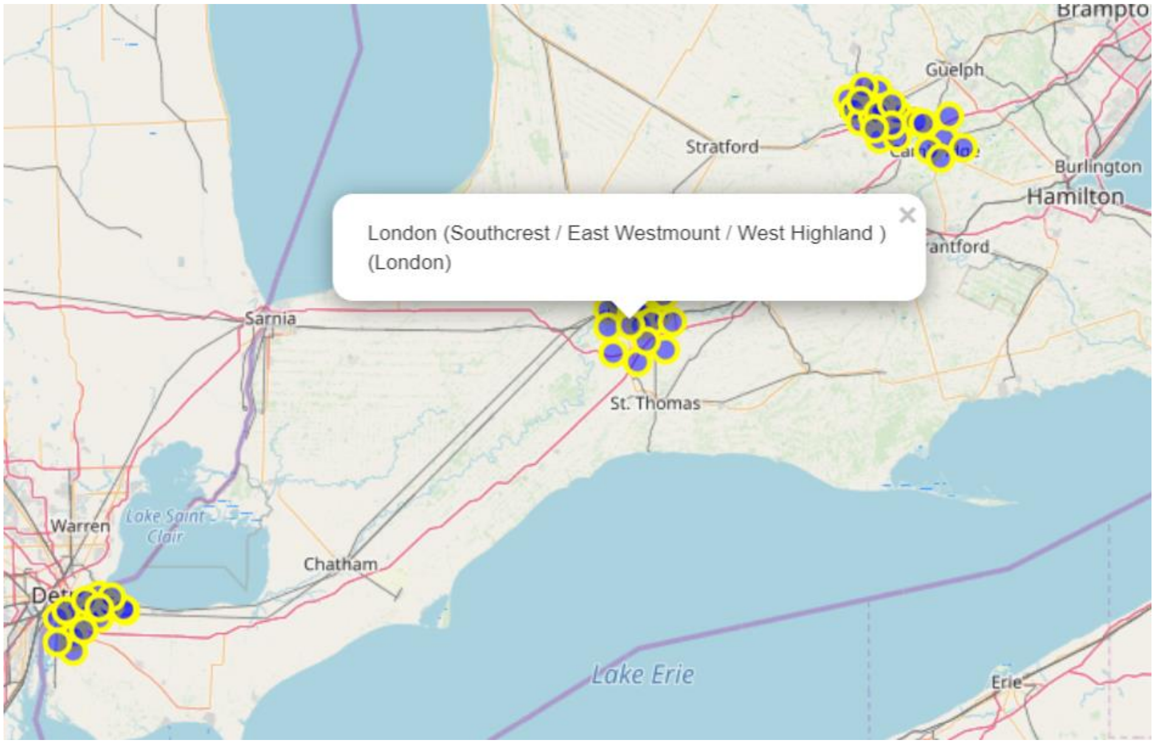
NORTH YORK LAWRENCE	75
NICHOLSON’S NOFRILLS TORONTO	52
ANTHONY'S NOFRILLS TORONTO	38
DAVE & CHARLOTTE'S NOFRILLS TORONTO	35
...	...
FORTINOS WATERDOWN HAMILTON ST	1
FRANCOIS' NOFRILLS TORONTO	1
MIKE & LORI'S NOFRILLS SCARBOROUGH	1
LANGSTAFF	1
LEONETTI'S NOFRILLS NORTH YORK	1

Table 7: Store Venues Categories

Venue Category	Number of Venues
Coffee Shop	60
Pizza Place	42
Café	27
Park	26
Sandwich Place	26
Grocery Store	24
Bakery	24
Restaurant	24
Fast Food Restaurant	24
Pharmacy	20

2.3.2 ROI Location Data

Same thing is done using FourSquare for the neighborhoods in the ROI. Figure 2.2 shows the geographical distributions of the neighborhoods in the ROI, and we can clearly see that they are distributed within the three boroughs – KWC, London, and Windsor.



(**Figure 2.2**: Geographical distributions of the Neighborhoods in ROI)

Table 8 shows the top and bottom five number of venues found for the neighborhoods in the ROI, and Table 9 shows the corresponding top 10 venue categories.

Table 8: ROI Neighborhood Venues Count

Neighborhood	Number of venues
Kitchener South Central	43
Kitchener Northwest	24
Waterloo Southeast	23
London Central	21
Waterloo South	16
…	…
Kitchener Southeast	1
London (East Tempo)	1
Tecumseh Outskirts (Windsor)	1
La Salle West (Windsor)	1
Cambridge Northwest	1

Table 9: ROI Neighborhood Venues Categories

Venue Category	Number of Venues
Coffee Shop	17
Pizza Place	10
Park	9

Restaurant	9
Convenience Store	9
Construction & Landscaping	8
Sandwich Place	8
Sporting Goods Shop	7
Pet Store	7
Café	6

Table 8 shows the top and bottom five number of venues found for the neighborhoods in the ROI, and Table 9 shows the corresponding top 10 venue categories.

3. Exploratory Data Analysis (EDA)

In this section, we will perform EDA on the venue categories we found around the stores and ROI locations. Since many of the default venue categories from FourSquare intrinsically belong to the same category, such as “Pizza Place”, “Sandwich Place” that are also restaurants, we will generalize the categories and condense them into fewer number of classes. After condensation, there are mainly 7 general venue classes – Entertainment, Functional, Restaurant, Service, Sports and Life, Stores, and Transportation. Table 10 shows some of the default venue categories inside the general venue classes. Notice that there are only two venue sub-categories falling inside the Functional class.

Table 10: Examples of Venue Categories inside the 7 General Venue Classes

Entertainment	Functional	Restaurant	Service	SportsNlife	Stores	Transportation
Entertainment	Distribution Center	Pizza Place	Bank	Park	Shop	Bus
Bar	Storage Facility	Café	Service	Gym	Store	Train
Pub	-	Sandwich Place	Lounge	Pool	Market	Station
Club	-	Breakfast Spot	Gas Station	Field	Grocery	Intersection
Tea Room	-	Burrito Place	Hotel	Garden	Pharmacy	Construction & Landscaping

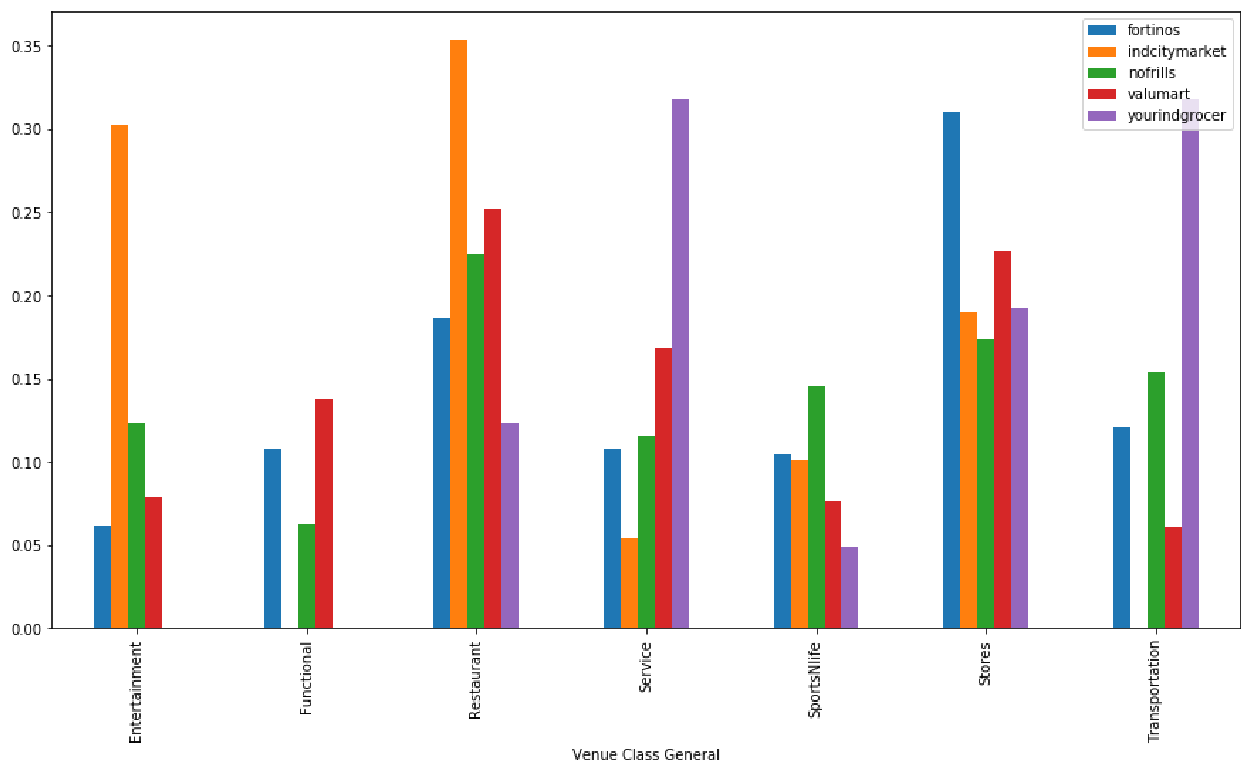
3.1 Store (Banner) Venues Analysis

After generalizing the venue categories into the 7 main classes for the stores, Table 11 shows the percentage of each main class for all the banners. The percentages are also plotted as in the bar chart and pie chart shown by Figure 3.1 and Figure 3.2, respectively.

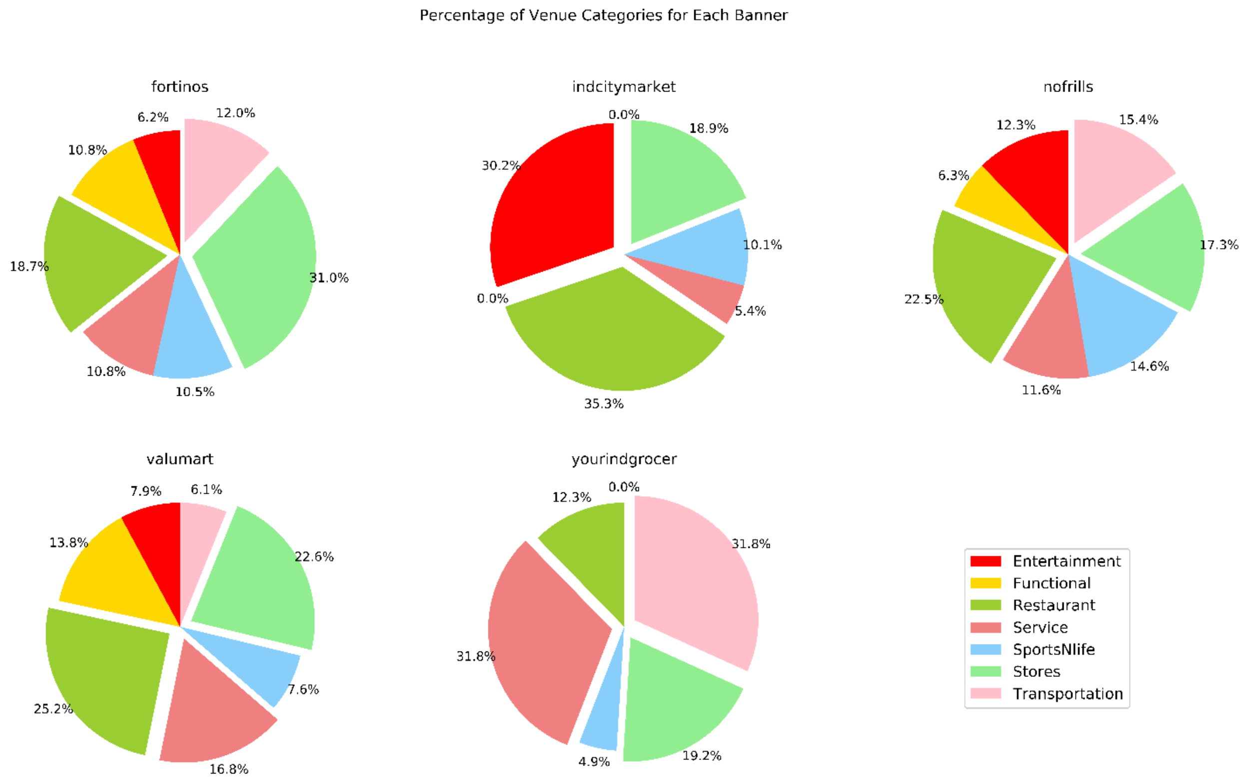
Table 11: Store (Banner) General Venue Class Percentage

	Fortinos	Independent City Market	No Frills	Valu-Mart	Your Independent Grocer
Entertainment	6.2%	30.2%	12.3%	7.9%	0.0%
Functional	10.8%	0.0%	6.3%	13.8%	0.0%
Restaurant	18.7%	35.3%	22.5%	25.2%	12.3%
Service	10.8%	5.4%	11.6%	16.8%	31.8%
SportsNlife	10.5%	10.1%	14.6%	7.6%	4.9%
Stores	31.0%	18.9%	17.3%	22.6%	19.2%
Transportation	12.0%	0.0%	15.4%	6.1%	31.8%

From the percentage table and the following bar chart, it is observed that 'Fortinos', 'No Frills', and 'Valu-Mart' have comprehensive venues while 'Independent City Market' and 'Your Independent Grocer' lack some surrounding venue types. Stores play an important role in having 'Fortinos' as well as 'Valu-Mart' in the area, and restaurants are the big factor for 'Independent City Market', 'No Frills', and 'Valu-Mart'. Entertainment is another popular neighbor for 'Independent City Market'. At last, 'Your Independent Grocer' seems to prefer service and transportation venues around it.



(**Figure 3.1**: Store (Banner) General Venue Class Percentage)



(**Figure 3.2**: Store (Banner) General Venue Class Percentage)

Table 12: Banners General Venue Classes in Order

	Fortinos	**Independent City Market**	**No Frills**	**Valu-Mart**	**Your Independent Grocer**
1	Stores	Restaurant	Restaurant	Restaurant	Transportation
2	Restaurant	Entertainment	Stores	Stores	Service
3	Transportation	Stores	Transportation	Service	Stores
4	Service	SportsNlife	SportsNlife	Functional	Restaurant
5	Functional	Service	Entertainment	Entertainment	SportsNlife
6	SportsNlife	Transportation	Service	SportsNlife	Functional
7	Entertainment	Functional	Functional	Transportation	Entertainment

Table 12 shows the general venue classes in order for each of the banners. From the pie charts in Figure 3.2 and the table above, it can be seen that Stores, Restaurant, and Transportation are the common top general venue categories among the banners. Therefore, when developing classification models, these three categories should be treated with more weights than the others.

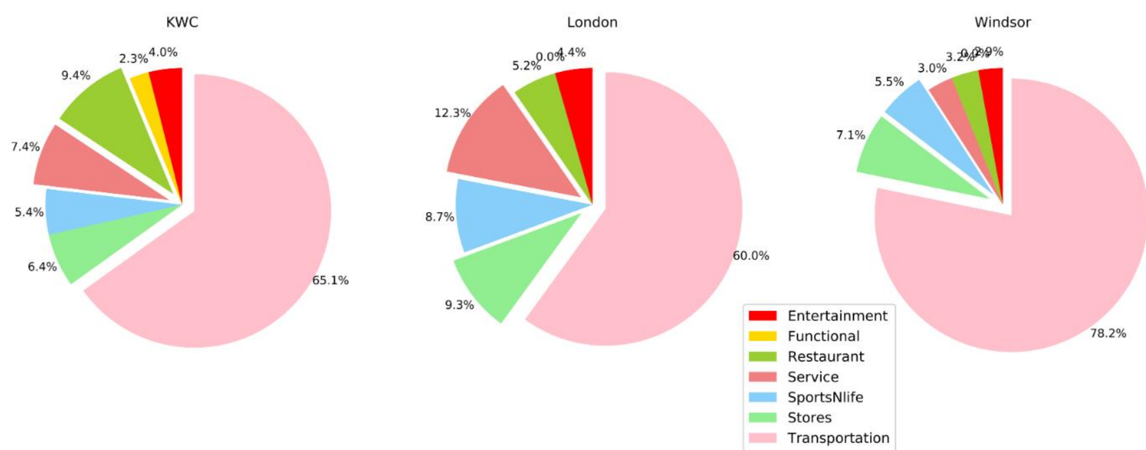
3.2 ROI Venues Analysis

The same venue category condensation is applied on the ROI venues, and Table 13 shows the result of percentage of the main categories for each borough in the ROI.

Table 13: ROI General Venue Class Percentage

	KWC	London	Windsor
Entertainment	4.0%	4.4%	2.9%
Functional	2.3%	0.0%	0.0%
Restaurant	9.4%	5.2%	3.2%
Service	7.4%	12.3%	3.0%
SportsNlife	5.4%	8.7%	5.5%
Stores	6.4%	9.3%	7.1%
Transportation	65.1%	60.0%	78.2%

Figure 3.3 below shows the percentages in pie charts. It is clearly observed that Transportation plays the most critical role in all the boroughs in the ROI. London and Windsor lack the functional type, which means they do not have distribution centers or storage facilities near the points of interest. Based on this information so far, 'Your Independent Grocer' seems to be the best fit for the boroughs in the ROI since its most preferred venue type is Transportation plus that Function is not very important for it. 'Fortinos' and 'No Frills' can be good alternatives because they also put Transportation as their top nearby venues and Functional in the bottom, at the meantime, their preference of Stores and Restaurants venues shows promising accordance to the boroughs in the ROI.



(**Figure 3.3**: ROI General Venue Class Percentage)

4. Modelling

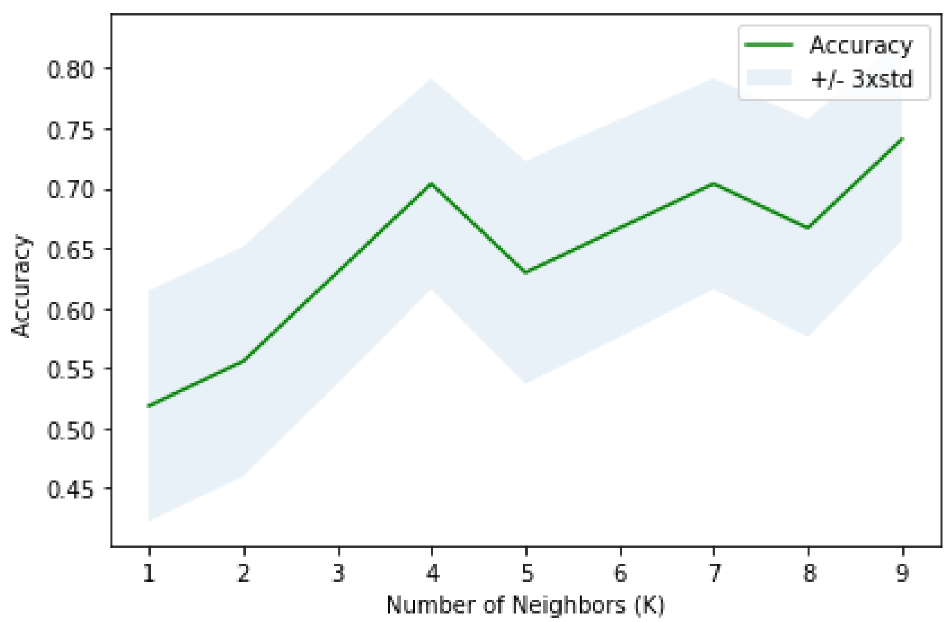
4.1 Classification Models to Predict an Appropriate Banner Based on General Venue Class Distribution

We have seen from the Section 3.1 that 'Stores', 'Restaurant', 'Transportation', and 'Entertainment' categories play important roles in having supermarket businesses nearby. Therefore, any neighborhoods that have none of these four categories around will be removed from our list.

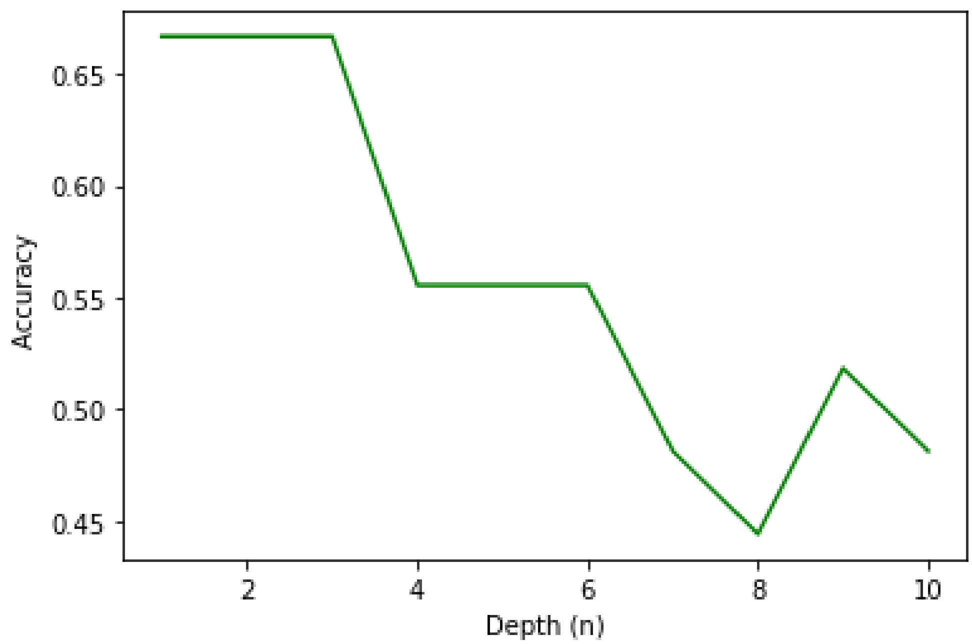
Four classification modelling methods are chosen for the purpose of predicting an appropriate banner by providing the model with the distribution of general venue categories. The four models are K-nearest Neighbors (KNN), Decision Tree, Support Vector Machine (SVM), and Logistic Regression. Evaluation test will be performed for each method, and the best method(s) will be selected to be used for the prediction based on the rank of their evaluation scores.

The complete data, which is used to derive Table 11, without being grouped by the banner types are separated into training and test sets with a distribution of 70% and 30%.

For KNN and Decision Tree models, tunings of the number of neighbors (k) and depth value (n) are performed. Figure 4.1 and 4.2, respectively, indicate the relationship of accuracy with k and n for these two models. From the accuracy trends, k of 7 and n of 3 are selected for KNN and Decision Tree models, respectively.



(**Figure 4.1**: Test of K's for KNN Model)



(**Figure 4.2**: Test of Depth Value for Decision Tree Model)

Table 14 shows the accuracy scores using different classification algorithms, and it indicates that KNN has the best performance overall. Therefore, KNN will be used as the prior predicting method to give an appropriate banner based on the venue information. The other methods can be used as backup at the same time, especially the Logistic Regression model that can provide additional information about probability.

Table 14: Evaluation Scores for Each Classification Method

Algorithm	Jaccard	F1-score	LogLoss
KNN	0.704	0.608	N/A
Decision Tree	0.667	0.586	N/A
SVM	0.704	0.605	N/A
Logistic Regression	0.667	0.533	1.518

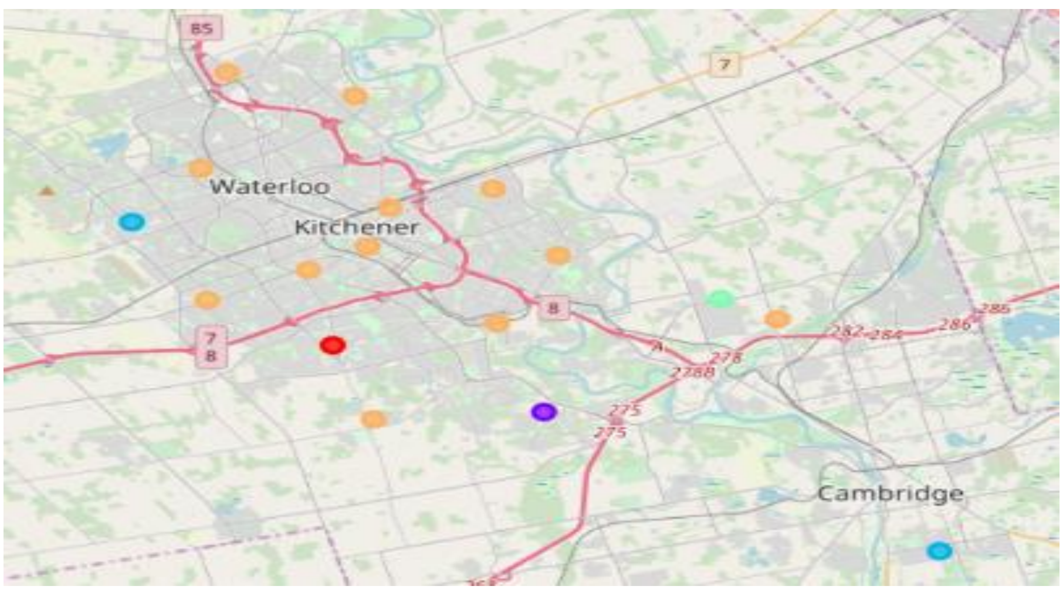
4.3 Clustering Model for Neighborhoods in the ROI

Before applying the classification model derived from the previous section directly on the neighborhoods in the ROI, I first developed a clustering model for the neighborhoods to further characterize them and get better understanding of the differences. In the study, 5 clusters are chosen to best separate these neighborhoods in the ROI. Table 14 shows the distribution of general venue classes for each cluster. We can see that Cluster 1-3 focus too much on the Transportation type, and Cluster 0 lacks necessary Functional and

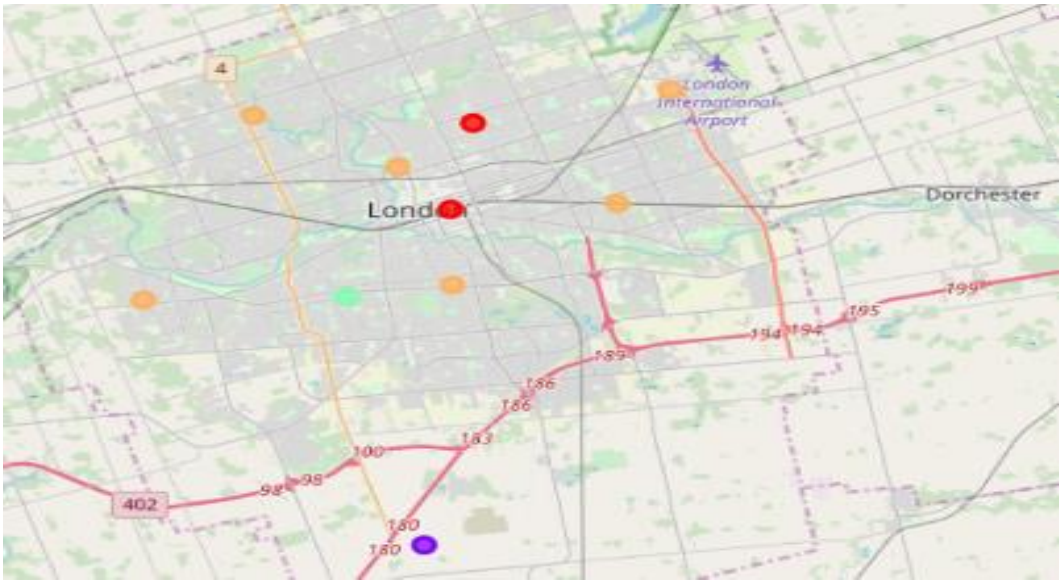
Transportation venue types useful for supermarkets. So, the only cluster that contains comprehensive venues with evenly distributed general venue types is Cluster 4.

Table 14: Neighborhood Clusters General Venue Class Percentage

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Entertainment	2%	0%	0%	4%	14%
Functional	0%	0%	0%	0%	5%
Restaurant	5%	0%	3%	1%	25%
Service	67%	0%	11%	4%	2%
SportsNlife	11%	0%	0%	5%	22%
Stores	15%	0%	0%	5%	26%
Transportation	0%	100%	86%	81%	7%



(**Figure 4.3**: KWC Neighborhood Clusters)



(**Figure 4.4**: London Neighborhood Clusters)

(**Figure 4.5**: Windsor Neighborhood Clusters)

4.4 Apply the Banner Classification Model on the Neighborhoods in ROI

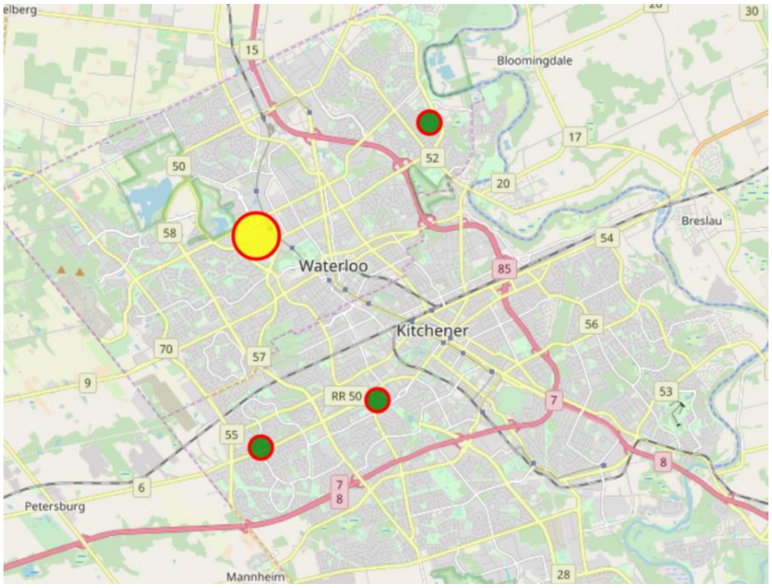
From last section, we found out that only neighborhoods in Cluster 4 suit to have a supermarket inside the area, so we will apply the classification model on these neighborhoods only. Table 15 shows the neighborhoods with suggested banners and general venue distribution. Our last mission is to propose to the private investor with a few of neighborhoods that have better venue distribution, which is critical for a supermarket's long-termly healthy business.

Table 15: Neighborhoods in Cluster 4 with Suggested Banners and Venue Distribution

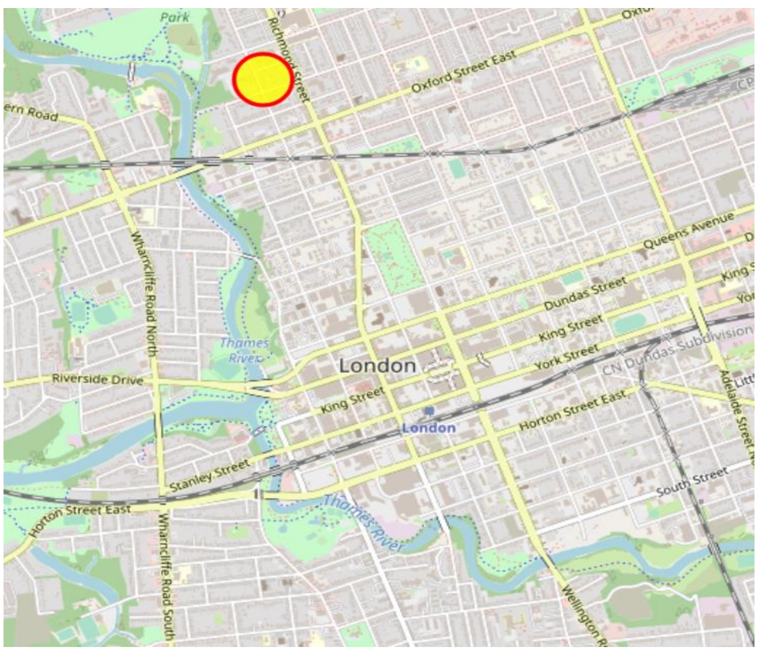
Neighborhood	Borough	Banner	Ent	Func	Rest	Serv	S.N.L	Stores	Trans
Cambridge Northwest	KWC	nofrills	0%	0%	0%	0%	0%	100%	0%
Kitchener Central	KWC	nofrills	0%	0%	69%	0%	0%	31%	0%
Kitchener East	KWC	nofrills	0%	0%	100%	0%	0%	0%	0%
Kitchener North Central	KWC	fortinos	0%	0%	50%	0%	41%	9%	0%
Kitchener Northeast	KWC	nofrills	0%	0%	100%	0%	0%	0%	0%
Kitchener Northwest	KWC	nofrills	0%	0%	**36%**	**38%**	**5%**	**20%**	0%
Kitchener South	KWC	nofrills	60%	0%	0%	0%	22%	19%	0%
Kitchener South Central	KWC	nofrills	11%	0%	38%	0%	4%	46%	0%
Kitchener West	KWC	fortinos	0%	0%	**25%**	0%	**41%**	**34%**	0%
London (Riverbend / Woodhull / North Sharon Creek / Byron / West Westmount)	London	nofrills	0%	0%	0%	0%	70%	30%	0%
London (Sunningdale / West Masonville / Medway / NE Hyde Park / East Fox Hollow)	London	nofrills	0%	0%	23%	0%	77%	0%	0%
London (YXU / North and East Argyle / East Huron Heights)	London	nofrills	0%	0%	0%	0%	0%	100%	0%
London East (SW Argyle / Hamilton Road)	London	nofrills	0%	0%	0%	0%	44%	56%	0%
London North (UWO)	London	nofrills	**22%**	0%	**40%**	0%	**24%**	**14%**	0%
London South (East Highland / North White Oaks / North Westminster)	London	nofrills	0%	0%	59%	0%	0%	41%	0%
Waterloo East	KWC	fortinos	0%	0%	**25%**	**0%**	**41%**	**34%**	0%
Waterloo South	KWC	nofrills	**31%**	0%	**2%**	0%	**6%**	**8%**	**54%**
Waterloo Southeast	KWC	nofrills	0%	61%	22%	0%	3%	14%	0%
Windsor (East Forest Glade)	Windsor	nofrills	66%	0%	0%	0%	24%	10%	0%
Windsor (East Riverside)	Windsor	nofrills	0%	0%	0%	0%	0%	100%	0%
Windsor (Roseland)	Windsor	nofrills	0%	0%	27%	0%	0%	73%	0%
Windsor (University / South Cameron)	Windsor	nofrills	0%	0%	74%	0%	0%	26%	0%
Windsor (West Forest Glade / East Fontainbleu)	Windsor	nofrills	0%	0%	42%	0%	0%	58%	0%
Windsor East (East Walkerville)	Windsor	nofrills	0%	0%	0%	0%	90%	10%	0%

After inspecting the table, I have found totally 5 options for investing a supermarket including 2 outstanding (highlighted in yellow) and 3 fair (highlighted in green) ones. Four of them are in Kitchener-Waterloo, and the other one is in London. Their locations can be seen on the maps by Figure 4.6 and 4.7. My pick is based on selecting neighborhoods that have good venue type distributions, especially restaurant, store, and transportation as discussed in Section 3.1.

After searching by Google map, there is no similar store like 'No Frills' or 'Fortinos' by Loblaws that offers duplicated functions. However, similar supermarkets do exist in these neighborhoods to drive competition. Discussion on these is beyond the scope of this analysis report.



(**Figure 4.6**: Proposed store locations in Kitchener-Waterloo)



(**Figure 4.7**: Proposed store locations in London)

5. Conclusion

In this analysis, I collected venue data for the selected Loblaw store locations in western Ontario and neighborhoods in Kitchener-Waterloo-Cambridge, London, and Windsor regions. The venue data for the stores is used to study uniqueness of the banners, which are provided by Loblaw to private investors to join its supermarket franchise. The venue data for the neighborhoods is used to understand the differences among them and also used as features to select appropriate store banners within the area. I built classification models using the transformed venue data of the banners, and used the best model to predict appropriate banners for the neighborhoods in the regions of interest. I also applied clustering model to group the neighborhoods using their transformed venue data and picked good ones for proposing investment. At the end, five store locations with appropriate banners were proposed by using the classification and clustering models with a few exploratory data analyses on the locations.

6. Future Work

The classification models were able to achieve ~70% accuracy in the classification of the banners. However, the store (banner) location data contains a problem that the number of stores for certain banners is too small comparing to others. For example, No Frills and Fortinos have many more stores than those of 'Independent City Market' and 'Your Independent Grocer'. This could cause bias when analyzing the venue profiles. If possible, more store locations should be collected for these banners even if outside of Ontario.

Models and analysis in this study mainly focused on geographical data. However, when proposing stores and locations investment, many other factors should be considered. The other factors contain but are not limited to demographics, competitors, average return on investment around the regions of interest. These additional factors should be taken into consideration to provide optimized and comprehensive investment advice.