

DATA SCIENCE PROFESSIONAL CERTIFICATE FINAL CAPSTONE PROJECT

Analysis and Modelling of Loblaw Banners Selection for Private Investors

Yiteng Liang

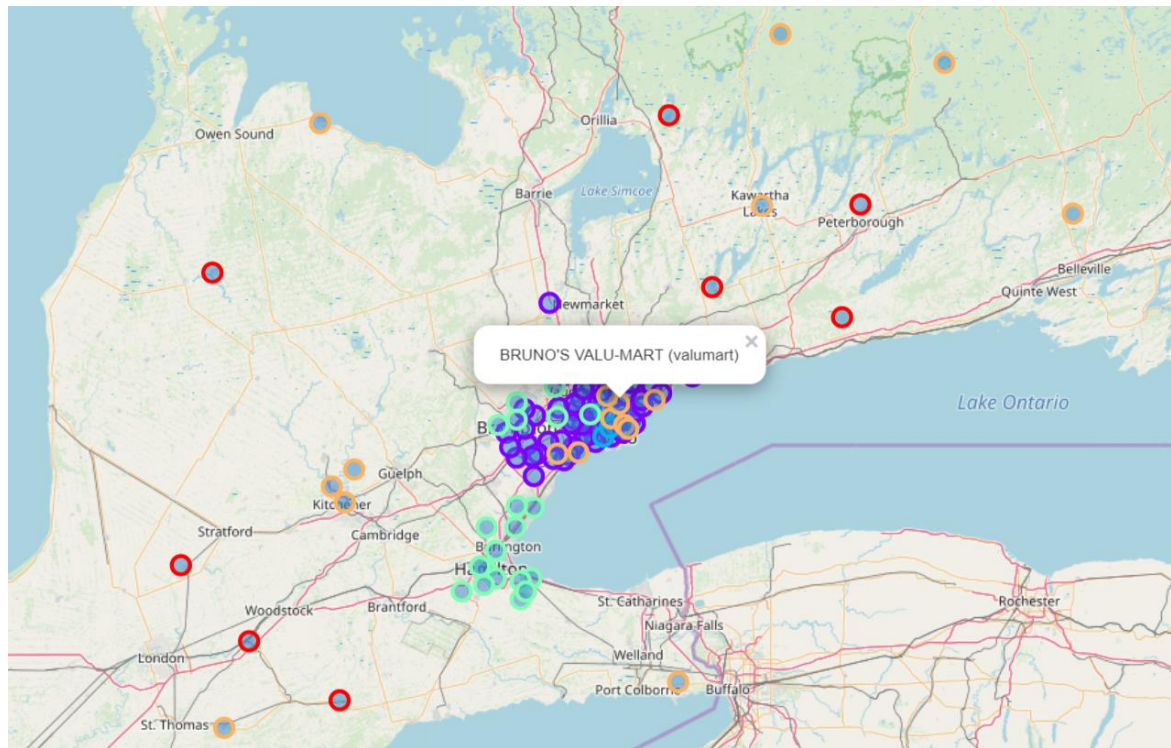
March 22, 2020

Business Problem

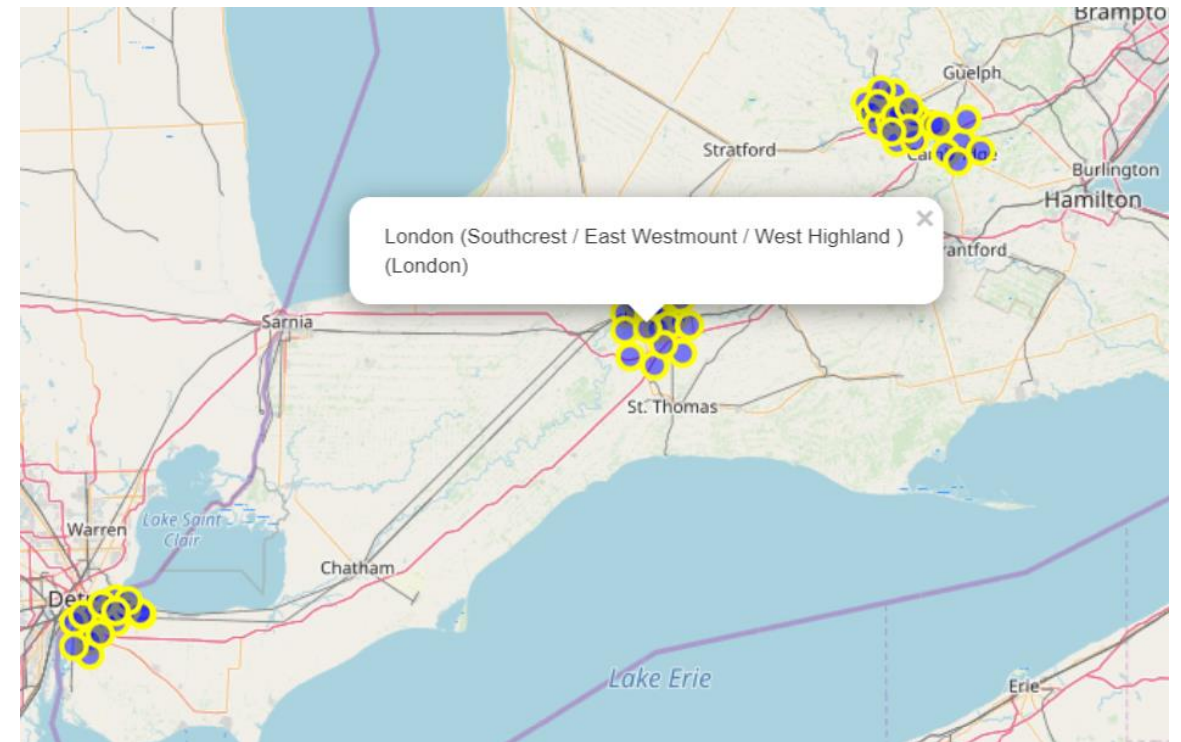
- A private investor wants to open a grocery store or supermarket through joining the Loblaw franchises in Ontario.
- Available franchise banners from Loblaw that are available for selection include.
 - [No Frills™](#)
 - [Independent City Market™](#)
 - [Fortinos™](#)
 - [Provigo™](#)
 - [Valu-Mart™](#)
 - [Your Independent Grocer™](#)
- Pick a location where one of the banners from the list could be the best fit.
- Expected locations should be in the regions of interest (ROI), including Kitchener-Waterloo-Cambridge (KWC), London, and Windsor.
- Help to get clear and detailed insight of the locations and neighborhood features around them.

Data Preparation - Locations

- Addresses and coordinates for the existing stores with the banners in Ontario.
- Addresses and coordinates for the neighborhoods in the ROI.



(Existing stores in Ontario with the banners)



(Neighborhoods in the ROI)

Data Preparation – Venue Data

- Venue data through FourSquare API is collected for the coordinates of stores and neighborhoods.
- Venue categories are further generalized into 7 main classes - Entertainment, Functional, Restaurant, Service, Sports and Life, Stores, and Transportation.

Top 10 Store Venue 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

Top 10 Neighborhood Venue 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

Exploratory Data Analysis (EDA)

- EDA on the 7 generalized venue class distributions for the existing stores.
 - Stores, Restaurant, and Transportation are the common top general venue categories among the banners.

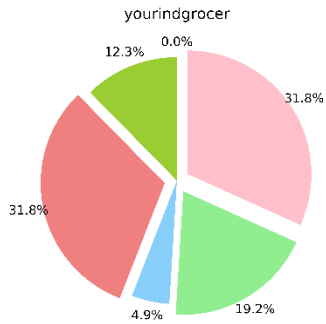
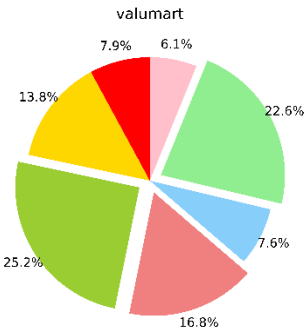
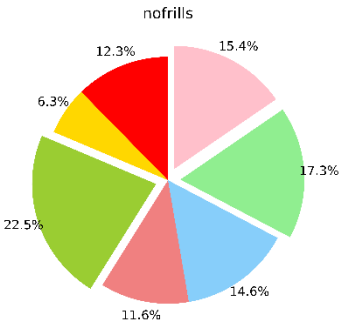
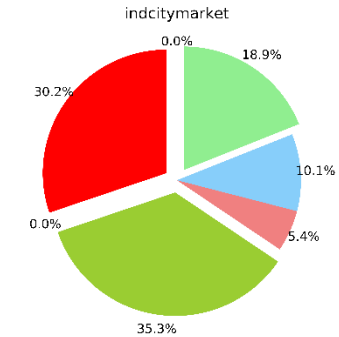
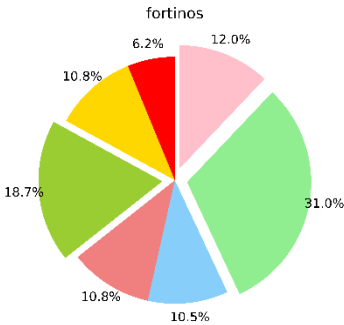
Store (Banner) General Venue Class Percentage

	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%

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

Percentage of Venue Categories for Each Banner



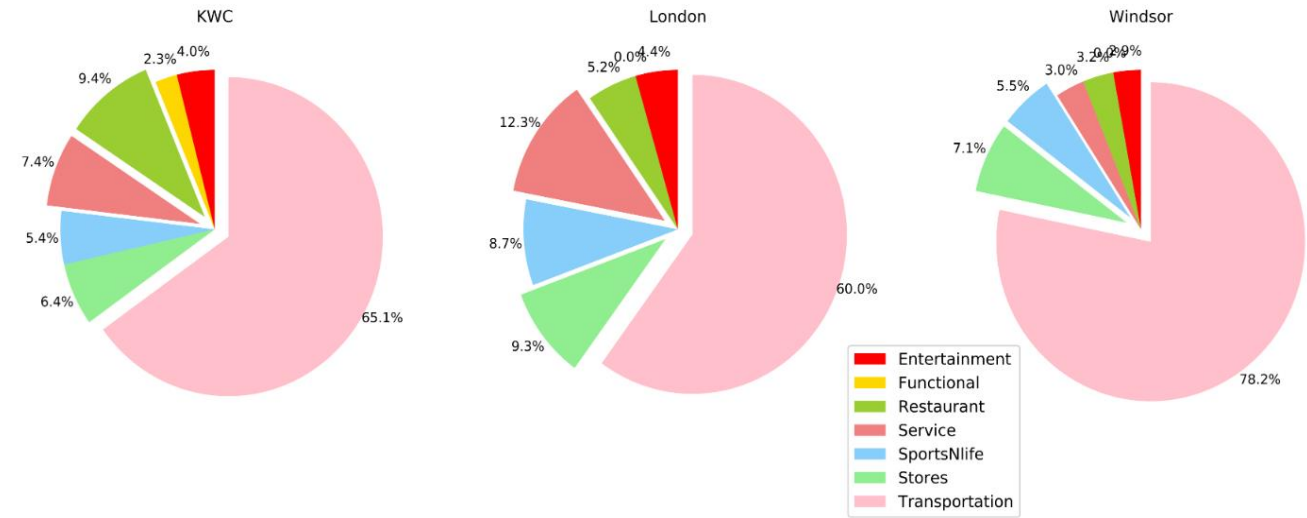
(Store (Banner) General Venue Class Percentage)

Exploratory Data Analysis (EDA)

- EDA on the 7 generalized venue class distributions for the neighborhoods in the ROI.
 - Transportation plays the most critical role in all the boroughs of the ROI.
 - London and Windsor lack the functional type.

ROI General Venue Class Percentage

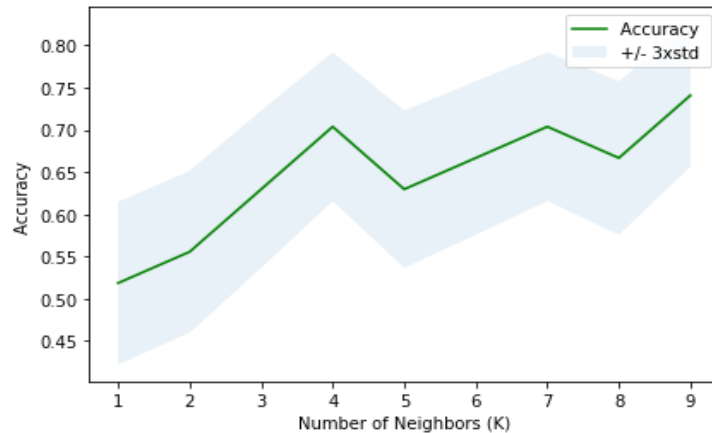
	KWC	London	Windsor
Entertainment	4.0%	4.4%	2.9%
Functional	2.3%	0.0%	0.0%
Restaurant	9.4%	5.2%	3.2%
Service	7.4%	12.3%	3.0%
SportsNlife	5.4%	8.7%	5.5%
Stores	6.4%	9.3%	7.1%
Transportation	65.1%	60.0%	78.2%



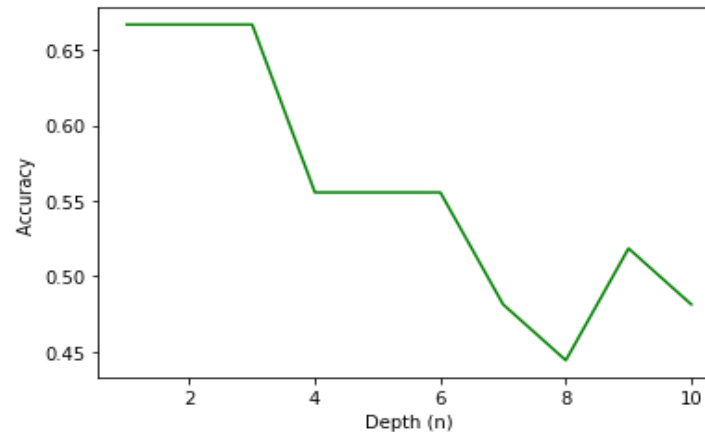
(ROI General Venue Class Percentage)

Classification Model

- K-nearest Neighbors (KNN), Decision Tree, Support Vector Machine (SVM), and Logistic Regression Transportation classification models are used to predict banners based on the input of venue general class distributions.
- Data is separated into 70% and 30% used for training and testing respectively.
- Tunings of the number of neighbors (k) and depth value (n) are performed for KNN and Decision Tree models.
- **KNN model has the best performance**



(Test of K's for KNN Model)



(Test of Depth Value for Decision Tree Model)

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

Clustering Model

- Neighborhoods in the ROI are clustered and insights are found by understanding the differences among the clusters.
- Cluster 1-3 focus too much on the Transportation type, and Cluster 0 lacks necessary Functional and Transportation venue types useful for supermarkets.
- The only cluster that contains comprehensive venues with evenly distributed general venue types is Cluster 4.

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%

Proposed Store Locations and Banners

- 5 options for investing a supermarket including 2 outstanding (highlighted in yellow) and 3 fair ones.
- Four of them are in Kitchener-Waterloo, and the other one is in London.
- My pick is based on selecting neighborhoods that have good venue type distributions, especially restaurant, store, and transportation.

Proposed Store Locations and Banners

Neighborhood	Borough	Banner
Kitchener Northwest	KWC	No Frills
Kitchener West	KWC	Fortinos
London North (UWO)	London	No Frills
Waterloo East	KWC	Fortinos
Waterloo South	KWC	No Frills

Conclusion and Future Work

- Collected venue data for the selected Loblaw store locations in western Ontario and neighborhoods in Kitchener-Waterloo-Cambridge, London, and Windsor regions.
- Built classification models using the transformed venue data of the banners and predict appropriate banners for the neighborhoods in the ROI.
- Applied clustering model to group the neighborhoods and picked good ones for proposing store locations and banners.
- Five store locations with appropriate banners were proposed.
- If possible, more store locations should be collected for the banners with fewer existing stores than others in Ontario
- When proposing stores and locations investment, many other factors should be considered - demographics, competitors, average return on investment. These should be taken into consideration to provide optimized and comprehensive investment advice.