

Local Yarn Stores are the New Black

Mollie E. Conrad, MSc.

October 15, 2019

Abstract

This analysis was conducted in fulfillment of the IBM Data Science Certification Capstone Project requirement.

In this study, the Foursquare API is used to determine which location in the Region of Waterloo is the most viable for a “Local Yarn Store” (LYS). LYSs are emerging as a niche spot for young to middle-aged adults.

Geographical data from Foursquare is used to determine the *best* location for a new LYS, should a prospective business-owner wish to know. This new location will likely be one that isn’t too close to other LYSs, or any local big box stores that are likely to sell similar products for lower costs. Additionally, it would be ideal to place the new location near high traffic, high population density parts of the Region.

Five locations are suggested based on the data analysis: 1 in Waterloo, 2 in Kitchener, and 2 in Cambridge. Of the 5, the Kitchener locations are suggested to be the most viable for success based on the chosen criteria.

Please see:

https://nbviewer.jupyter.org/github/meconrad/Coursera_Capstone/blob/master/YarnStoresAreTheNewBlack_CapstoneWeek5.ipynb
for Jupyter Notebook.

Chapter 1

Introduction

1.1 What *is* a ‘Local Yarn Store’, Anyway?

“Local Yarn Stores” (LYS) seem to be a niche establishment to an unknowing individual, used only by grannies and crazy cat ladies. But what many people don’t know is that within the knitting and crochet fibre community, MANY young folx are ditching the big box stores like ”Michael’s” for unique and inspiring LYSs. It is here that you can find yarn hand-dyed by your super talented neighbour, yarn hand-spun from fleece, sourced from the next town over, and sometimes even hand-brewed specialty coffee (Hello, Knitting Loft, Toronto)! LYSs are seriously underestimated treasure troves.

1.2 Business Problem Outline

Here’s the big idea: we want to open up a *new* LYS to serve the yarn-obsessed community.

Currently in KW, there are only 3 LYS *that I know of* within a 20 - 30 minute drive radius. We can use geographical data from Foursquare to determine any other LYS, as well as high traffic shopping locations. In addition to the Foursquare API, we can utilize publically available population data from Stats Canada to determine the Region of Waterloo municipal population density, and publically available municipal boundary data from the Region of Waterloo for visual mapping purposes. The LYS/ Shopping Center data, population data, and municipal boundary data will allow us to determine the most ideal location for a new LYS.

For the purposes of this project, we will assume we don’t already know the quantity and locations of *any* LYS within the Region of Waterloo.

The data utilized in this analysis will be provided in Chapter 2, the data analysis methodology will be described in Chapter 3, and a Discussion and Conclusions will be provided in Chapter 4.

1.3 Interest

Prospective LYS owners would be interested in these results. If looking into a different region, along with obtaining that Region’s specific GeoJSON and population data, one could simply alter the coordinate input. This code could also be useful for someone looking to open any other type of store; one would simply have to alter the Foursquare search queries. This code is really quite versatile and could easily be manipulated to serve the user’s needs.

Chapter 2

Data

2.1 Foursquare API: LYS, Michael's, and Shopping Center Locations

The Foursquare API was used to collect locations and information about LYS, Michael's stores, and shopping centers in the Region of Waterloo.

When querying for LYS, the search query “yarn” was used with a radius of 30.0 km and a threshold limit of 30 results. The JSON results were transformed into a pandas dataframe with from `pandas.io.json.json_normalize`. In terms of cleaning the results, only location columns were kept (along with the venue name and category), the venue category was changed to only state the category name, and only the last term was kept when the column name contained 2 or more terms. Based on prior knowledge of the region, venue results that either were returned by mistake or were previously closed were dropped. Any missing venue entries (i.e Shall We Knit?) were added manually. The final dataframe is given in Figure 2.1 for example.

	name	categories	lat	lng	distance	cc	country	address	postalCode	city	state	crossStreet
1	Galt House Of Yarn	Knitting Store	43.3581	-80.3175	17819	CA	Canada	110-7 Grand Ave S	N1S 2L3	Cambridge	ON	NaN
3	All Strung Out Fine Yarns	Arts & Crafts Store	43.5456	-80.2507	22266	CA	Canada	36 Quebec St	N1H 2T4	Guelph	ON	Baker St
4	Yarnbird	Knitting Store	43.6808	-80.4301	25902	CA	Canada	22 Mill St W	N0B 1S0	Elora	ON	NaN
5	Shall We Knit?	Knitting Store	43.4652	-80.5181	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Figure 2.1: Pandas dataframe used within the Jupyter notebook to tabulate Waterloo Region LYS.

When querying for Michael's Arts and Crafts stores, the search query “Michaels” was used with a radius of 30.0 km and a threshold limit of 30 results. The cleaning methods mentioned above were used. Any entries that were not in the category “Arts and Crafts” were dropped.

When querying for Shopping Malls, the search query “shopping mall” was used with a radius of 50.0 km and a threshold limit of 50 results. The threshold and radius were increased to allow for clustered shopping areas to be evident. The same cleaning methods were used. Bus lines and duplicate venues were dropped; i.e., stores within malls that contained the mall name.

2.2 Stats Canada 2016 Census Population Data

Population data for the KW region can be obtained from StatsCanada:

https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page_Hierarchy-Hierarchie.cfm?Lang=E&Tab=1&Geo1=CMACA&Code1=541&Geo2=PR&Code2=35&SearchText=Kitchener%20-%20Cambridge%20-%20Waterloo&SearchType=Begins&SearchPR=01&B1=Population&TABID=1&type=0

From this URL, I obtained 2016 census data for the following municipalities:

- Kitchener
- Cambridge
- Waterloo
- North Dumfries
- Wilmot
- Woolwich
- Wellesley

Once the .CSV files were read in with pandas, they were cleaned in such a way as to reduce confusion and to ensure compatibility across Data. An example of one of these dataframes can be found in Table 2.1. To get to this cleaned stage, multiple columns were dropped (columns = ['Note', 'Flag_Total', 'Flag_Male', 'Flag_Female', 'Flag_Total.1', 'Flag_Male.1', 'Flag_Female.1'] - these are simply notes provided within the table), and only rows pertaining to 'Population and dwellings' and 'Age characteristics' were kept. In hindsight, 'Age characteristics' could have been dropped as well, although it was thought to potentially be useful at the time.

Table 2.1: Municipality of Waterloo: 2016 Census Data

Topic	Characteristics	Total	Male	Female	Total_ON	Male_ON	Female_ON
1	Population and dwellings	104986			13448494.0		
2	Population and dwellings	98780			12851821.0		
3	Population and dwellings	6.3			4.6		
4	Population and dwellings	46096			5598391.0		
5	Population and dwellings	40381			5169174.0		
6	Population and dwellings	1639.8			14.8		
7	Population and dwellings	64.02			908699.33		
8	Age characteristics	104985	52490.0	52500.0	13448495.0	6559390.0	6889105.0
9	Age characteristics	16445	8440.0	8005.0	2207970.0	1131615.0	1076350.0
10	Age characteristics	4740	2440.0	2300.0	697360.0	357300.0	340055.0
11	Age characteristics	5590	2900.0	2685.0	756085.0	387320.0	368765.0
12	Age characteristics	6120	3100.0	3020.0	754530.0	387000.0	367530.0
13	Age characteristics	73375	37030.0	36340.0	8988865.0	4412115.0	4576750.0
14	Age characteristics	7440	3875.0	3560.0	811670.0	416730.0	394945.0
15	Age characteristics	11700	6285.0	5420.0	894390.0	457815.0	436570.0
16	Age characteristics	7620	4175.0	3435.0	874350.0	436345.0	438010.0
17	Age characteristics	6150	3110.0	3040.0	864765.0	420800.0	443965.0
18	Age characteristics	5950	2955.0	2995.0	842485.0	402900.0	439585.0
19	Age characteristics	6580	3155.0	3425.0	871875.0	418465.0	453415.0
20	Age characteristics	7415	3505.0	3905.0	937385.0	453105.0	484275.0
21	Age characteristics	7795	3780.0	4020.0	1056345.0	517170.0	539175.0
22	Age characteristics	7020	3485.0	3535.0	989460.0	482280.0	507180.0
23	Age characteristics	5705	2700.0	3005.0	846145.0	406505.0	439640.0
24	Age characteristics	15165	7015.0	8155.0	2251655.0	1015655.0	1236000.0
25	Age characteristics	4905	2350.0	2555.0	737745.0	352475.0	385270.0
26	Age characteristics	3460	1625.0	1840.0	528645.0	250455.0	278185.0
27	Age characteristics	2550	1155.0	1395.0	393010.0	179810.0	213195.0
28	Age characteristics	1980	890.0	1085.0	291185.0	126825.0	164360.0
29	Age characteristics	2270	995.0	1275.0	301075.0	106090.0	194985.0
30	Age characteristics	1415	570.0	845.0	188470.0	73045.0	115425.0
31	Age characteristics	780	400.0	380.0	88390.0	27400.0	60990.0
32	Age characteristics	65	20.0	45.0	21210.0	5150.0	16060.0
33	Age characteristics	10	0.0	10.0	3005.0	495.0	2520.0
34	Age characteristics	100.0	100.0	100.0	100.0	100.0	100.0
35	Age characteristics	15.7	16.1	15.2	16.4	17.3	15.6
36	Age characteristics	69.9	70.5	69.2	66.8	67.3	66.4
37	Age characteristics	14.4	13.4	15.5	16.7	15.5	17.9
38	Age characteristics	2.2	1.9	2.4	2.2	1.6	2.8
39	Age characteristics	39.0	38.0	40.0	41.0	40.0	41.9
40	Age characteristics	37.7	35.6	39.7	41.3	40.2	42.4

The population density of each municipality was extracted from their respective dataframes using `pd.iloc()`. The extracted values can be found in Table 2.2. These population densities were placed in a new dataframe, along with the Municipality names and their respective coordinates, in order to be easily mapped with *Folium*; this Python package allows the population density to be visually presented as a layer on a map.

Table 2.2: Population Density of Each Municipality within the Region of Waterloo

Municipality	Population Density
Waterloo	1639.8
Kitchener	1705.2
Cambridge	1149.6
North Dumfries	54.5
Wilmot	77.9
Woolwich	76.7
Wellesley	40.5

2.3 Region of Waterloo Municipal Boundary Data

The Region of Waterloo publically provides files regarding municipal boundaries. A GeoJSON file and .csv file of the Region of Waterloo municipality boundaries were obtained from

<https://open-kitchenergis.opendata.arcgis.com/datasets/RMW::regional-boundaries/data>.

GeoJSON files are given below:

GeoJSON (Waterloo Region)

https://opendata.arcgis.com/datasets/dc4eff944b774abdb6ee0e1931a8663f_17.geojson

GeoJSON (Municipalities)

https://opendata.arcgis.com/datasets/2840815b1dff4989b8c8513541a00b49_0.geojson

Table 2.3: Region of Waterloo Municipality Boundary Data - from .CSV file

	OBJECTID	MUNICIPALITY	PERIMETER	MUN_AREA	MUN_ID	CATEGORY
0	1	WILMOT	69.83794803	266.18335094	104	MUNICIPALITY
1	2	WATERLOO	41.25743596	65.23647683	100	MUNICIPALITY
2	4	WOOLWICH	100.70451526	329.68346085	103	MUNICIPALITY
3	5	NORTH DUMFRIES	85.62600126	190.22873175	101	MUNICIPALITY
4	6	CAMBRIDGE	60.04553772	115.36289433	106	MUNICIPALITY
5	7	WELLESLEY	65.45636003	278.41574333	102	MUNICIPALITY
6	1608	KITCHENER	59.75282052	138.40565649	105	MUNICIPALITY

Table 2.4: Joined Dataframe: Municipality Boundaries and Coordinates

	Municipality	Lat	Long	Population Density	OBJECTID	PERIMETER	MUN_AREA	MUN_ID	CATEGORY
0	WILMOT	43.435	-80.628	77.9	1	69.83794803	266.18335094	104	MUNICIPALITY
1	WATERLOO	43.464	-80.52	1639.8	2	41.25743596	65.23647683	100	MUNICIPALITY
2	WOOLWICH	43.534	-80.387	76.7	4	100.70451526	329.68346085	103	MUNICIPALITY
3	NORTH DUMFRIES	43.314	-80.391	54.5	5	85.62600126	190.22873175	101	MUNICIPALITY
4	CAMBRIDGE	43.432	-80.31	1149.6	6	60.04553772	115.36289433	106	MUNICIPALITY
5	WELLESLEY	43.5322	-80.7658	40.5	7	65.45636003	278.41574333	102	MUNICIPALITY
6	KITCHENER	43.453	-80.495	1705.2	1608	59.75282052	138.40565649	105	MUNICIPALITY

Chapter 3

Analysis

Once the data from Foursquare, Stats Canada, and the Region of Waterloo have been cleaned and prepared, they can be used to generate interactive maps of the Region with the Python package *Folium*. These maps will act as the foundation for our conclusions.

Before mapping with *Folium*, the population densities acquired from Stats Canada were plotted in a bar-graph with *Matplotlib* (see Figure 3.1). From this, it is evident that the 3 municipalities with the highest population density are: Kitchener, Waterloo, and Cambridge. This is no surprise.

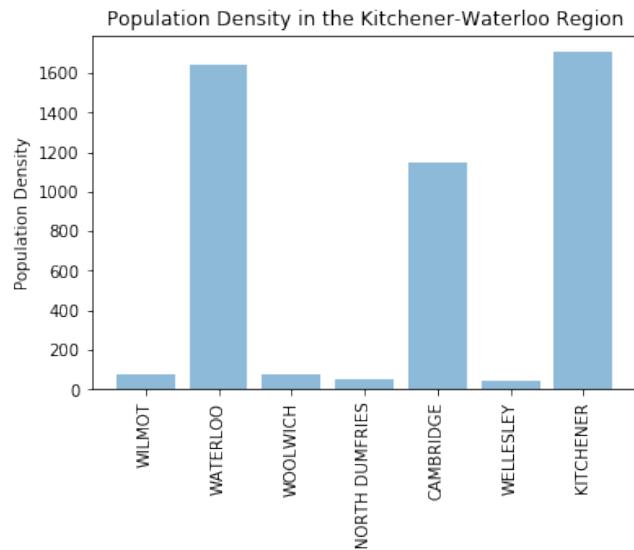


Figure 3.1: Region of Waterloo Municipality Population Density

Municipal boundaries obtained from the Region of Waterloo can be found mapped in Figure 3.2; municipal center coordinates (red), LYS locations (blue), and Michael's locations (green) are mapped in Figure 3.3. High traffic shopping centers (like malls; yellow) are added in Figure 3.4. The ideal location will be in high density locations (dark blue), near highways, near the yellow markers (shopping centers), and away from green and blue markers (LYS and Michael's).

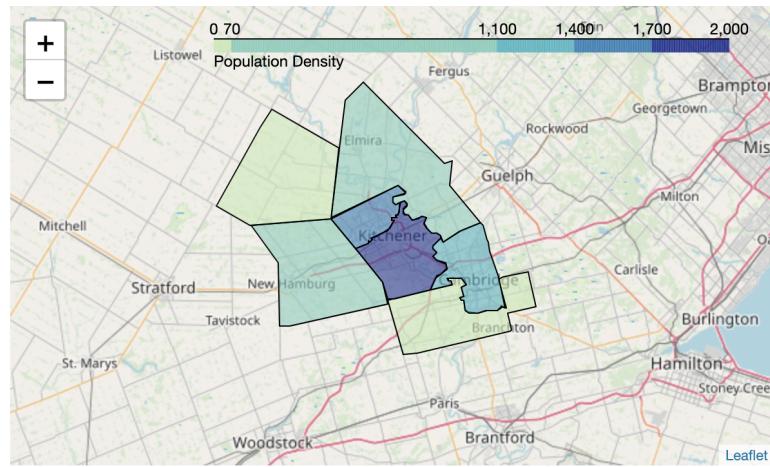


Figure 3.2: Municipal Boundaries Plotted with Folium (Python), considering population density (see legend).

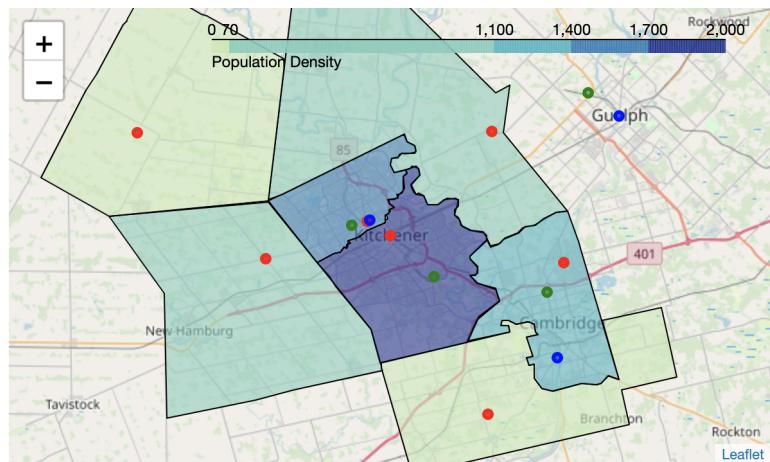


Figure 3.3: Municipal boundaries mapped with municipal center coordinates(red), LYS locations (blue), and Michael’s locations (green). Population density also mapped (see legend).

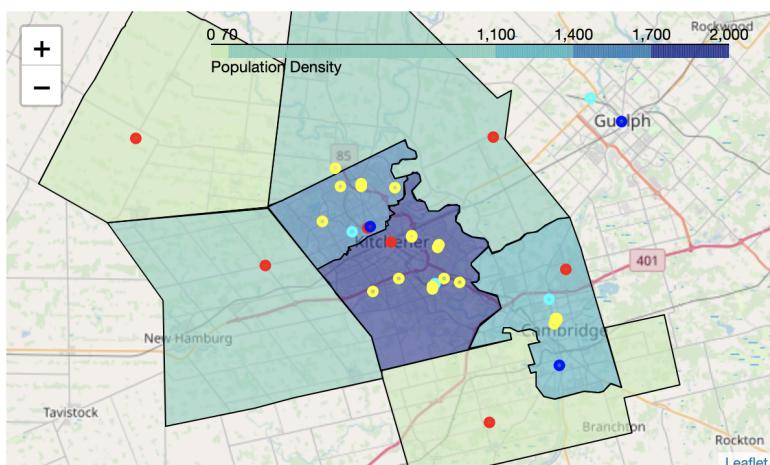


Figure 3.4: Municipal boundaries mapped with municipal center coordinates (red), LYS locations (blue), Michael’s locations (cyan), and population density (see legend).

Chapter 4

Results & Discussion

Local Yarn Stores (LYS) are unique little havens for both young and elderly knitters and crocheters. We are interested in knowing where the most ideal location would be for a new LYS, in terms of both success and accessibility.

By inspection of the *Folium* maps, let's choose 5 locations that meet the following criteria:

1. Near high population density regions
2. Far enough away from other yarn stores that competition isn't a concern
3. Near highways for easy access
4. Perhaps near high traffic shopping centers

The geographical coordinates of the 5 potential locations are provided in Table 4.1 and are mapped in Figures 4.1 and 4.2. The locations that are near a Michaels store would still be ideal; Michael's isn't exactly the same as a LYS, thus enabling each type of store to have their own niche markets. In fact, customers would likely view this location as a “two birds, one stone” scenario. Zoomed-in *Folium* maps of each individual location are given in Figures 4.3, 4.4, and 4.5; example building locations have been selected.

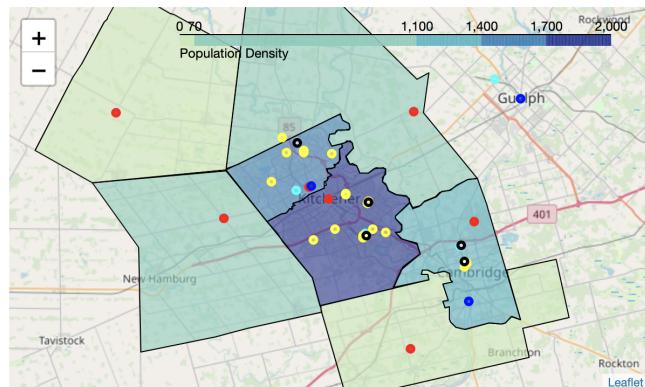


Figure 4.1: Final 5 potential new LYS positions (Black). Municipal boundaries mapped with municipal center coordinates (red), LYS locations (blue), Michael's locations (cyan), shopping centers (yellow), and population density (see legend).

Table 4.1: 5 Ideal Locations for a New LYS. HWY = closest highway to the location, Lat = latitude, Long = longitude.

Location	Local Shopping	HWY	Near LYS?	Near Michaels?	Population Density	Lat	Long
Waterloo	Conestoga Mall	85	No	No	1639.8	43.5049	-80.5356
Kitchener	Stanely Park Mall	8, 7	No	No	1705.2	43.4498	-80.4452
Kitchener	Manitou Mall, Best Mall	8	No	Yes	1705.2	43.4194	-80.4476
Cambridge	Cambridge Smart Centers - Hespeler/ Pinebush	401, 24	No	Yes	1149.6	43.4103	-80.3261
Cambridge	Cambridge Center Mall	401, 24	No	No	1149.6	43.3949	-80.3229

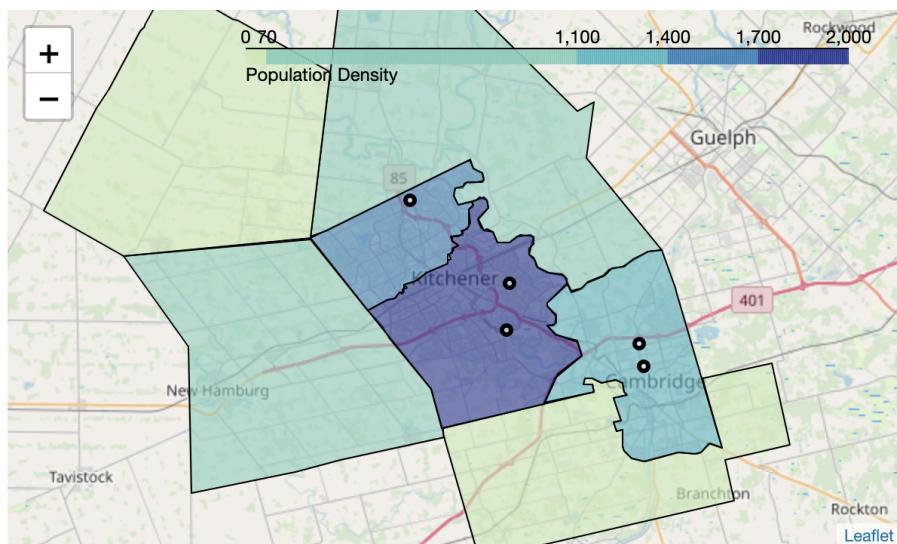


Figure 4.2: Final 5 potential new LYS positions (circles). Municipal boundaries mapped with population density (see legend).



Figure 4.3: Ideal LYS location in Waterloo, ON: near Conestoga Mall.

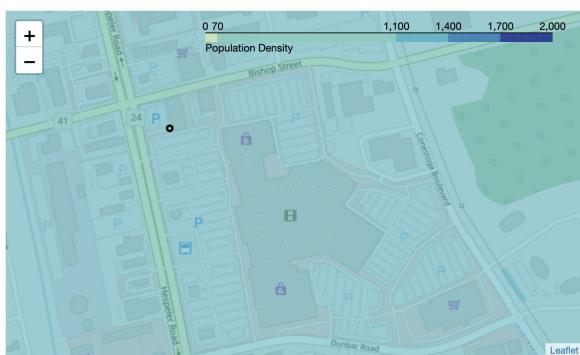


(a) Ideal LYS location: near Stanley Park Mall.

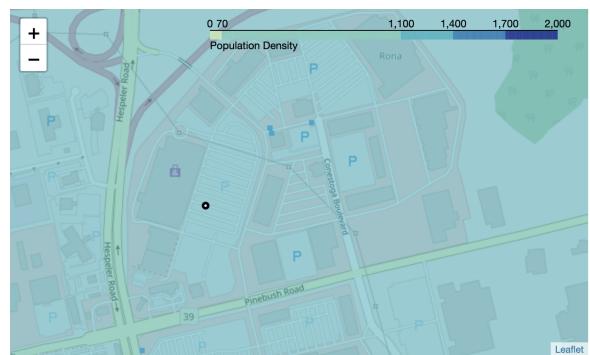


(b) Ideal LYS location: near Fairway Rd.

Figure 4.4: Kitchener Ideal LYS Locations



(a) Ideal LYS location: near Cambridge Center Mall.



(b) Ideal LYS location: near Cambridge Smart Centers

Figure 4.5: Cambridge Ideal LYS Locations

Chapter 5

Conclusions and Future Work

Considering the population density, the central nature within the region, and the nearness to local highways, the potential Local Yarn Store (LYS) locations in KITCHENER would likely be the most successful.

In the future, one could further consider the population data obtained from the Stats Canada 2016 Census. This data supplies the population within specific age groups and genders. While, in my opinion as a 27-year-old woman, the yarn obsession is real for humans of all ages, LYS do tend to be more expensive than, for example, Michael's. One skein of hand-dyed yarn runs at approximately \$32 CAD while one skein from Lion Brand Yarn, sold at Michael's, runs at approximately \$6 – \$10 CAD. Bearing this in mind, established adults might be more keen on spending nearly triple the amount. Prioritizing municipalities where the density of adults aged 30+ is maximized may help boost the success of the new LYS.

Please see:

https://nbviewer.jupyter.org/github/meconrad/Coursera_Capstone/blob/master/YarnStoresAreTheNewBlack_CapstoneWeek5.ipynb
for Jupyter Notebook.