Capstone Project - The Battle of Neighborhoods

Opening a new Artisanal Coffee Shop in Toronto

Table of Contents

Introduction

Background

Toronto is the most populous city in Canada with a developed economy that is highly diversified with strengths in technology, design, financial services, life sciences, education, arts, fashion, aerospace, environmental innovation, food services, and tourism. This presents a host of opportunities and challenges for an entrepreneur.

While the urban setting, large population and strong economy drive strong demand, there is also significant competition and one of the highest operating cost for any location in Canada. Hence leveraging data science and the insights it produces in identifying ideal conditions and testing hypotheses provides a better chance of success for any new business.

Problem Description

A (hypothetical) client wishes to open a new artisanal coffee shop, which will be their foray into the Canadian market. They plan to offer a cozy and comfortable space for both office workers and casual visitors to sit down, while also offering a take-out option. There will be a strong emphasis on high quality coffee at a reasonable cost.

One of the most important success drivers for a coffee shop business is its location and hence they are looking to leverage data to identify an ideal Toronto neighborhood that will broadly satisfy the following criteria:

- A strong demand – they feel that they test well with office/business workers (white collar) and may also work well with college students (colleges in vicinity a plus, but not mandatory)
- A robust population – the preference is for a neighborhood with a large population (especially working age), with ideally a high average per capita earning
- Limited competition – the location should have a low density of coffee shops, but not so low that it indicates a lack of viability. Something in the middle is preferable

Target Audience

The target audience for the location report is the hypothetical coffee shop's management team. There is significant emphasis within the organization on data driven decision making, which prompted this project.

They expect to see a report with a recommendation for a neighborhood that meet their criteria, the data behind it, data source, methodology, and any further actions/caveats.

Data

Outline

To achieve the objective set out by the client to identify an ideal neighborhood (that meet their criteria) for their coffee shop:

i) A reliable data source needs to be identified for Toronto neighborhoods
ii) Leverage Foursquare for businesses and competition data (coffee shops and cafes)
iii) Use techniques like k-means clustering to identify and cluster similar neighborhoods
iv) Use demographics data to make a recommendation on an ideal neighborhood from a candidate cluster

Data sources and feature selection

The following data sources will be used to produce all the insights that will lead to a reliable recommendation.

1. Toronto Open Data Portal (https://open.toronto.ca/)
    a. Neighborhoods (https://open.toronto.ca/dataset/neighbourhoods/)
        i. Data extract: CSV directly into pandas DataFrame
        ii. Data:
            1. Neighborhood name
            2. Neighborhood latitude
            3. Neighborhood longitude

	Neighborhood	Longitude	Latitude
0	Wychwood	-79.425515	43.676919
1	Yonge-Eglinton	-79.403590	43.704689
2	Yonge-St.Clair	-79.397871	43.687859
3	York University Heights	-79.488883	43.765736
4	Yorkdale-Glen Park	-79.457108	43.714672

    b. Neighborhood Profiles (https://open.toronto.ca/dataset/neighbourhood-profiles/)
        i. Data Extract: neighbourhood-profiles-2016-csv directly into pandas DataFrame
        ii. Data:
            1. Neighborhood name
            2. Average income of resident by neighborhood (census data)
            3. Total count of working age population by neighborhood (census data)

Neighborhood	Income	Work Age Pop
Agincourt North	30,414	11,305
Agincourt South-Malvern West	31,825	9,965
Alderwood	47,709	5,220
Annex	112,766	15,040
Banbury-Don Mills	67,757	10,810

c. Foursquare API
    i. Data: Coffee Shops, Cafes, Offices/Businesses and Colleges within each neighborhood in 1 km radius
        1. Neighborhood name
        2. Neighborhood latitude
        3. Neighborhood longitude
        4. Venue
        5. Venue latitude
        6. Venue longitude
        7. Venue category

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wychwood	43.676919	-79.425515	CocoaLatte	43.681768	-79.425158	Café
1	Wychwood	43.676919	-79.425515	Baker and Scone	43.681614	-79.426075	Café
2	Wychwood	43.676919	-79.425515	Contra Cafe	43.669107	-79.426105	Café
3	Wychwood	43.676919	-79.425515	Krave Coffee	43.680740	-79.429417	Coffee Shop
4	Wychwood	43.676919	-79.425515	Starbucks	43.671530	-79.421400	Coffee Shop

Methodology

Data preparation and cleansing

The following steps were performed to extract data and make it useful for further analysis.

1. The two data sets (Neighborhood and Neighborhood Profiles) from Toronto Open Data were extracted into pandas DataFrames using read_csv()
2. Between the two datasets, I had all the base data that I needed to do my analysis. This included coordinates, its working age population and average income for all 140 official Toronto neighborhoods. Additional columns in the dataset, which were dropped as they were not relevant in this case.
3. Although both datasets were from the same source, there were multiple discrepancies between the datasets on neighborhood names, which was used to join the datasets. The first issue was that the

names did not match between datasets – the Neighborhoods dataset had area short code included in the name. The second issue was that 4 names had small differences between the two datasets. The first issue was fixed using split() function and the second was resolved using inplace updates at row level

AREA_SHC	AREA_LON	AREA_NAI	AREA_DESC	LONGITUDE	LATITUDE
94	94	Wychwoo	Wychwood (94)	-79.42551495	43.67691927
100	100	Yonge-Egli	Yonge-Eglinton (100)	-79.40359017	43.70468937
97	97	Yonge-St.(	Yonge-St.Clair (97)	-79.39787077	43.68785887
27	27	York Unive	York University Heights (27)	-79.48888293	43.76573649
31	31	Yorkdale-(	Yorkdale-Glen Park (31)	-79.45710807	43.71467153

```
# Fixing four value mismatches in Neighborhood data
torODneigh['Neighborhood'].replace({'Cabbagetown-South St.James Town':'Cabbagetown-South St. James Town',
                                    'Mimico':'Mimico (includes Humber Bay Shores)',
                                    'North St.James Town':'North St. James Town',
                                    'Weston-Pellam Park':'Weston-Pelham Park'},inplace=True)
```

4. The latitude and longitude data were used to query Foursquare venue explore functionality to identify (a maximum of 100 venues per category) venues in each neighborhood within relevant categories (Coffee Shops/Cafes, Colleges and Offices)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wychwood	43.676919	-79.425515	CocoaLatte	43.681768	-79.425158	Café
1	Wychwood	43.676919	-79.425515	Baker and Scone	43.681614	-79.426075	Café
2	Wychwood	43.676919	-79.425515	Contra Cafe	43.669107	-79.426105	Café
3	Wychwood	43.676919	-79.425515	Krave Coffee	43.680740	-79.429417	Coffee Shop
4	Wychwood	43.676919	-79.425515	Starbucks	43.671530	-79.421400	Coffee Shop

5. The resulting dataset from Foursquare was filtered into coordinates (only) for coffee shops and cafes to be used in heatmap generation. Similarly, office names (venues) and coordinates were extracted to add to map to provide visualization

	Venue Latitude	Venue Longitude
0	43.681768	-79.425158
1	43.681614	-79.426075
2	43.669107	-79.426105
3	43.680740	-79.429417
4	43.671530	-79.421400

	Venue	Venue Latitude	Venue Longitude
0	Camp Wahanowin Toronto Office	43.705018	-79.405610
1	LSBF Canada	43.706365	-79.400756
2	Woori Education & Immigration Consulting	43.706918	-79.400280
3	Royal LePage Condo Showroom	43.704663	-79.408436
4	Canadian Tire Home Office	43.704766	-79.398349

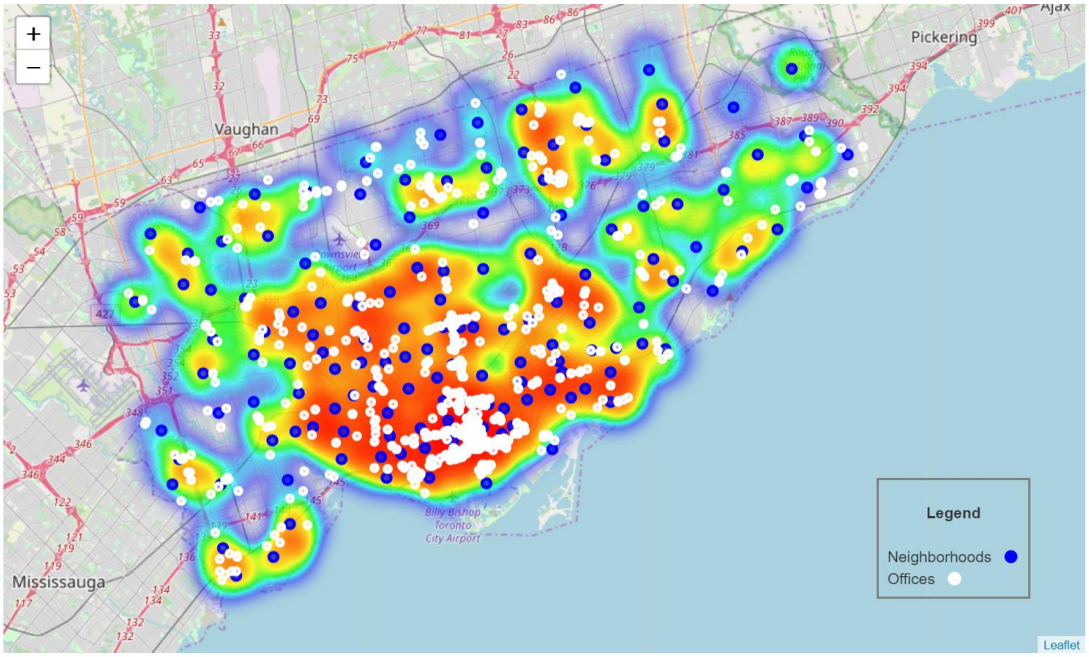
6. This provides all relevant data for further exploratory data analysis and visualizations to aid in decision making

Exploratory Data Analysis

The following steps were performed in order to make sense of the data that was collected:

1. Visualize the spread of coffee shops in Toronto with a heatmap with an overlay of neighborhoods and offices (potentially the main demand driver) for added context. As expected, downtown had the most

concentration of offices and coffee shops, but it also revealed that there were neighborhoods with offices/businesses that were relatively underserved for its demographics



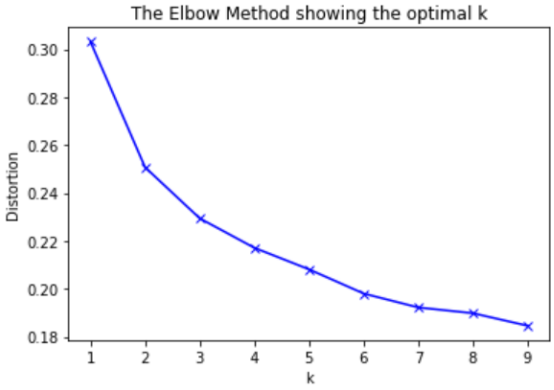
2. The next steps were to explore the Foursquare data further – which neighborhoods returned most venues, use one hot encoding to identify venue stats by neighborhood and grouping the neighborhoods by mean occurrence for each venue
3. Another important analysis was to identify the top 10 most popular venues in each neighborhood. This will give a good view on each neighborhood and will be the basis of decision making later on

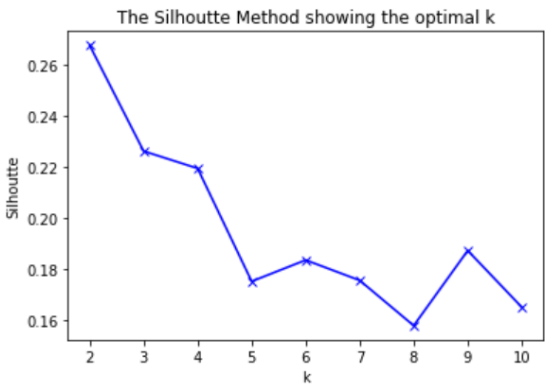
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt North	Coffee Shop	Office	College Academic Building	Bakery	Business Service	College Classroom	University	College Quad	College Engineering Building	College Football Field
1	Agincourt South-Malvern West	Office	Café	Coffee Shop	College Classroom	College Science Building	College Bookstore	College Academic Building	College Cafeteria	Student Center	College Theater
2	Alderwood	Coffee Shop	Office	Café	College Classroom	College Baseball Diamond	General College & University	College Cafeteria	Advertising Agency	Student Center	Community College
3	Annex	Café	Coffee Shop	Office	College Academic Building	College Library	General College & University	College Administrative Building	University	College Arts Building	College Residence Hall
4	Banbury-Don Mills	Café	Coffee Shop	Art Gallery	Office	Trade School	Cosmetics Shop	College Quad	College Engineering Building	College Football Field	College Gym

Clustering

Now that we have all the relevant datasets, its time to use K-Means clustering to identify neighborhood clusters that are similar in nature and will help narrow the search for an ideal neighborhood to open the artisanal coffee shop.

To identify the optimal K, both the elbow and silhouette methods were utilized. Based on the results, a K of 2 seemed ideal and hence used.





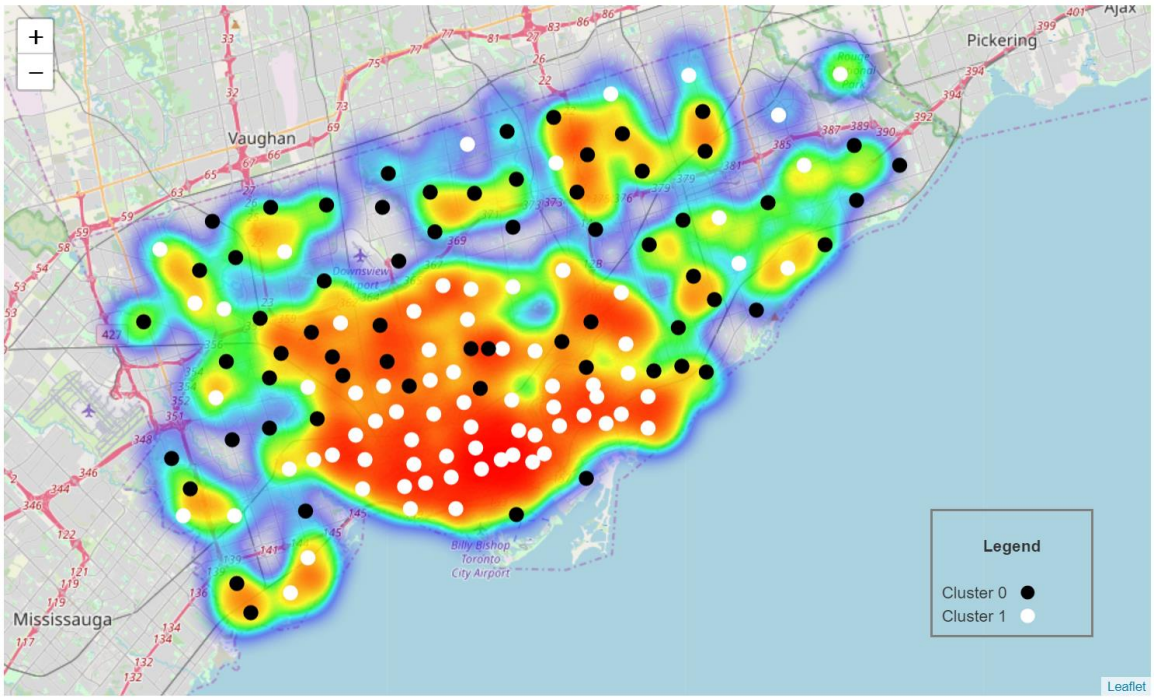
The resulting clusters show a clear distinction with one primarily having coffee shops/cafes as the top venue and the other cluster having offices as the most popular venue.

	index	Neighborhood	Longitude	Latitude	Income	Work Age Pop	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th M Comn Ve
0	0	Wychwood	-79.425515	43.676919	54,460	6,420	1	Café	Coffee Shop	American Restaurant	Dessert Shop	College Gym	College Rec Center	General College Universit
1	1	Yonge-Eglinton	-79.403590	43.704689	89,330	5,860	0	Office	Coffee Shop	Café	College Academic Building	Tech Startup	General College & University	Tea Roo
2	2	Yonge-St.Clair	-79.397871	43.687859	114,174	5,960	0	Office	Coffee Shop	Café	College Academic Building	Campaign Office	College Classroom	General College Universit
3	3	York University Heights	-79.488883	43.765736	29,958	12,290	0	Office	Coffee Shop	Café	Automotive Shop	Fraternity House	College Residence Hall	College Enginee Building
4	4	Yorkdale-Glen Park	-79.457108	43.714672	38,527	5,860	0	Office	Coffee Shop	Café	General College & University	Trade School	Medical Center	Medical School

Results

Neighborhood clusters

The resulting clusters, when visualized on a map with a heatmap for coffee shops included reveal the clear distinction between the two clusters. Cluster 0 is the clear choice as it meets the criteria of high concentration of offices/business (most popular venue), which supply the primary clientele for the coffee shop. Cluster 0 also show lower density of coffee shops, mostly situated in the fringe areas of high coffee shop density. This fares well with our criteria of low competition but with an indication of viability.



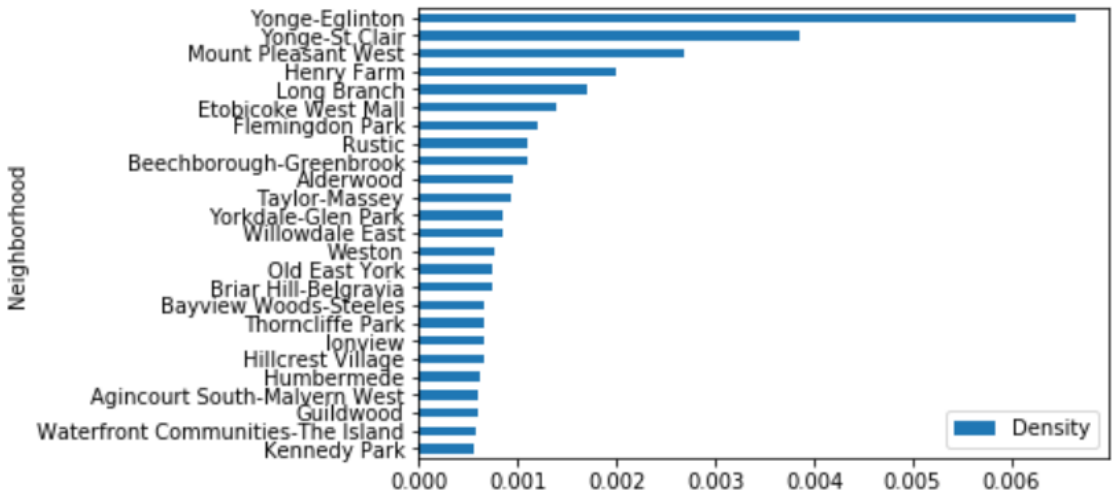
Picking the ideal neighborhood

Now that we have narrowed down to a cluster, it's time to explore one of the remaining two criteria (high average income and high working age population) to pick the most appropriate neighborhood to open the coffee shop.

To achieve a weighted score, let's leverage the income and population columns to build a score that's slightly biased towards higher income areas. Then the dataset is sorted for the highest score, pointing us towards the ideal candidate neighborhood: **Waterfront Communities-The Island** (**criteria 1 – strong demand as indicated by cluster and criteria 2 – robust population as indicated by weightage, met**).

	Neighborhood	Latitude	Income	Work Age Pop	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	
0	Waterfront Communities-The Island	43.633880	70,600	45,105	0	Office	Coffee Shop	Café	Tech Startup	College Academic Building	College Library	Genera & Unive
1	Yonge-St.Clair	43.687859	114,174	5,960	0	Office	Coffee Shop	Café	College Academic Building	Campaign Office	College Classroom	Genera & Unive
2	Willowdale East	43.770602	45,326	25,850	0	Office	Café	Coffee Shop	College Administrative Building	Business Service	College Classroom	Cowork Space
3	St.Andrew-Windfields	43.756246	100,516	6,800	0	Office	College Academic Building	University	College Quad	College Classroom	College Communications Building	College Engine Buildin
4	Edenbridge-Humber Valley	43.670886	101,551	5,940	0	Office	Coffee Shop	Real Estate Office	University	College Quad	College Classroom	College Commu Buildin
5	Mount Pleasant West	43.704435	57,039	17,100	0	Office	Coffee Shop	Café	College Academic	Tech Startup	General College & University	School

Let's also create a coffee shop density per person based on the population and coffee shop count in the neighborhood. This will be used to verify the third criteria of whether the competition is low in the area.



The coffee shop/café density is below average for a location such as this. This means that even though there is competition, its low and will not hamper the new artisanal coffee shop's chances in any way. The analysis also shows that most of the coffee shops/cafes in the area are national/international chains that are unlikely to directly compete with the second wave coffee shop format. (**criteria 3 – limited competition met**)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude	Venue Category
Venue						
AMPM Coffee	2	2	2	2	2	2
Aroma Espresso Bar	1	1	1	1	1	1
Boxcar Social	2	2	2	2	2	2
Cafe Supreme	2	2	2	2	2	2
Cocoberry Cafe	2	2	2	2	2	2
Harbour Bar	1	1	1	1	1	1
Lavazza Espression	2	2	2	2	2	2
Mos Mos	1	1	1	1	1	1
Starbucks	5	5	5	5	5	5
Tim Hortons	5	5	5	5	5	5
Treats	1	1	1	1	1	1
World Café at Harbourfront Centre	2	2	2	2	2	2

There is also community driven support for business (such as Waterfront Business Improvement Area - https://www.waterfrontbia.com/) in the area. Apart from all this, depending on exact location, this neighborhood has potential for visitors/tourists, increasing the client base. A picturesque setting within the neighborhood may also inspire people to sit down and have a coffee while taking in the views.

Discussion

Further considerations, caveats

While demand, population and competition were taken into consideration to make the recommendation, there are other critical factors that may also need to be considered like rent, available of the right property and prevailing market conditions (which were not part of this analysis). Another consideration may be the menu selection, which should be to the liking of Torontonians, which also requires further analysis.

Conclusion

The data indicates that **Waterfront Communities-The Island** is an ideal neighborhood to open an artisanal coffee shop in Toronto. The following criteria laid out by the client have been met with this location.

- Strong demand - The neighborhood has enough businesses/offices to support a loyal clientele plus this neighborhood features high foot traffic areas, which is another advantage
- A robust population - The neighborhood has 45,000+ working age residents with a high average income
- Limited competition - The area has a low density of coffee shops compared to similar neighborhoods, but not so low as to create any doubts about the viability of an artisanal coffee shop in this area

Overall, Waterfront Communities-The Island is an ideal location to open a new artisanal coffee shop that meets all of the client's criteria and some more. Perhaps, a good location within the neighborhood is Queen's Quay Terminal (marked below), which could be further explored.