COURSERA CAPSTONE

IBM Applied Data Science Capstone

Starting a Whole Distribution business in Toronto: Using Data Science to select what product(s) to supply

By: Brittany Frederick

August 2019

Introduction

Wholesale and distribution is a complex business, requiring smart management of supply chains, costs and pricing, and more. Done right, it's a lucrative line of work. So it's fitting that the industry is well-represented on the 2017 PROFIT 500 ranking of Canada's Fastest-Growing Companies. However, one of the most important factors to consider when starting a new wholesale distribution business is deciding what type of product to supply/distribute to retailers.

Business Problem

The objective of this project is to help an investor/business owner select the optimal type of product to supply/distribute within a potential wholesale distribution startup company. This project will aim to analyze Neighborhoods in the city of Toronto and their respective most frequently occurring type of venues. Chances of success when starting a wholesale distribution business will be higher if you select a retail segment that dominates the city (i.e. many potential clients).

Target Audience for this project

This project would be primarily useful to investors looking to invest in a startup wholesale distribution business in the city of Toronto. It will help them determine what type of product to supply and distribute to clients. For example, if at the end of this project if it is determined that 'Grocery Stores' is the most frequently occurring type of venue in the majority of the Neighborhoods in Toronto, it might be a good idea to start a Grocery Distributor company in Toronto.

Data

To solve this problem the following data will be used:

- List of neighborhoods in Toronto.
- Latitude and Longitude of the neighborhoods.
- Most commonly occurring venue category in each neighborhood.

Sources of data

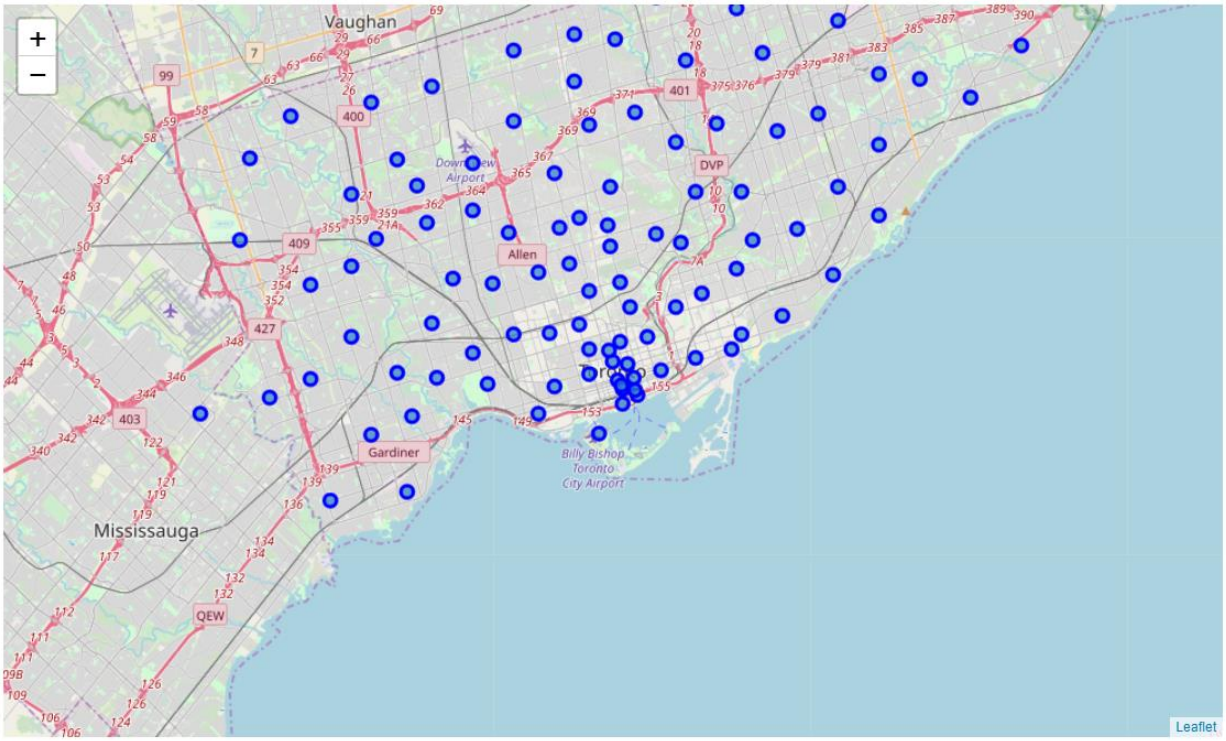
A list of neighborhoods will be extracted from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M using web scraping techniques via Beautiful Soup packages with Python. After a list of neighborhoods in Toronto with their respective Latitude and Longitude is established, the Foursquare API will be used to obtain venue categories for each neighborhood. It is these venue

categories that will provide insight for the best product to supply in the city of Toronto as a wholesale distributor.

Methodology

Initial steps involved scraping Toronto neighborhood data from Wikipedia using BeautifulSoup packages in Python. Data wrangling was then conducted, as there were 1) extra whitespace characters, 2) unnecessary header contents, and 3) Extra rows where a Borough in a Neighborhood was not named. Following this, data with no Borough names was dropped and data was grouped by postcode. Following all data wrangling, 103 unique Neighborhoods were identified.

Geographic data was then imported (provided by Coursera instructors). This data was used to find Latitude and Longitude coordinates for each Neighborhood. Folium in Python was then used to create an interactive map of all 103 Neighborhoods in Toronto.



This map will later be used for considerations of a potential location for a new Warehouse for this new Wholesale Distribution venture.

Following creation of the aforementioned map, the Foursquare API was used to find information about venues in each of the 103 Toronto neighborhoods:

- A call to each neighborhood using the Foursquare API was used to find all venues in a 500 mile radius of each neighborhood. This was stored in a dataframe.

- Exploratory data analysis was conducted to gain an initial perspective of the venue data. Stats were found such as: 2253 unique venues were found among all 103 Neighborhoods, 271 unique categories for types of venues were found among the 2253 unique venues.

After utilizing the Foursquare API to obtain venue data, a dataframe was created to display each neighborhood and the most frequently occurring types of venues for those neighborhoods.

	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	Adelaide, King, Richmond	Coffee Shop	Café	Steakhouse	Bar	Hotel	Cosmetics Shop	Restaurant	Thai Restaurant	Burger Joint	American Restaurant
1	Agincourt	Lounge	Sandwich Place	Breakfast Spot	Clothing Store	Chinese Restaurant	Drugstore	Discount Store	Dog Run	Doner Restaurant	Donut Shop
2	Agincourt North, L'Amoreaux East, Milliken, St...	Park	Playground	Coffee Shop	Women's Store	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant
3	Albion Gardens, Beaumond Heights, Humbergate, ...	Grocery Store	Liquor Store	Coffee Shop	Fast Food Restaurant	Beer Store	Sandwich Place	Fried Chicken Joint	Pizza Place	Pharmacy	Comic Shop
4	Alderwood, Long Branch	Pizza Place	Coffee Shop	Gym	Pharmacy	Skating Rink	Sandwich Place	Dance Studio	Pool	Pub	Women's Store

Using Pandas and Numpy packages in Python, the aforementioned dataframe was analyzed to determine how many times each unique category of venue appeared as the

‘1st Most Common Venue’.

```
neighborhoods_venues_sorted['1st Most Common Venue'].value_counts()
```

```
Coffee Shop               20
Park                      15
Café                       6
Grocery Store              5
Home Service               4
Trail                      3
Pizza Place                3
Fast Food Restaurant       2
Breakfast Spot             2
Bakery                     2
Furniture / Home Store     2
Indian Restaurant          2
Sandwich Place             2
Gym / Fitness Center       2
Lounge                     1
Gym                        1
Chinese Restaurant         1
French Restaurant          1
Motel                      1
Pet Store                  1
College Stadium            1
Food Truck                 1
Bar                        1
Pool                       1
Electronics Store          1
Restaurant                 1
Shopping Mall              1
Pharmacy                   1
Clothing Store             1
History Museum             1
Garden Center              1
Convenience Store          1
Beer Store                 1
Playground                 1
Department Store           1
Hotel                      1
Drugstore                  1
Sporting Goods Shop        1
Bank                       1
Greek Restaurant           1
Thai Restaurant            1
Airport Service            1
Auto Garage                1
Falafel Restaurant         1
Name: 1st Most Common Venue, dtype: int64
```

Results

Coffee Shops more often than any other type of venue was the 1st most common venue type for all Neighborhoods (Cafes were shortly behind in 3rd place for this spot). I.e., in 20 out of 103 Toronto Neighborhoods, Coffee Shops are the most common type of retailer. While this may not seem to be a high percentage, it is important to consider that all other types of venues that were found via the Foursquare API occur at lower frequencies than Coffee Shops. Due to these results, starting a Wholesale Coffee Supply business venture would be invaluable to the city of Toronto, with a high assurance that demand for Coffee Shop supplies would be high.

Discussion and Considerations

The next step for this project would be to determine the best location for this new warehouse. Using the Folium interactive map that was generated, it would be interesting to color all 20 of the Neighborhoods that have more Coffee Shops than any other type of venue and observe any potential clusters of these neighborhoods.

It is also important to consider potential ROI. This project does not take into account any potential ROI for supplying Coffee Shop supplies to the city of Toronto; it is entirely possible that supplying a different product pertinent to retailers that do not occur as frequently as Coffee Shops might provide a higher ROI.

Conclusion

The business problem at hand for this project was to determine what type of product to supply for a new Wholesale Distribution business in Toronto. Using Data Science it was determined that in all the Neighborhoods in Toronto, Coffee Shops were the most commonly occurring type of venues. Therefore, it would be wise to supply Coffee Shop supplies to the city of Toronto, as it is guaranteed there would be a high demand due to the prevalence of the Coffee industry in the city.

Sources

https://www.canadianbusiness.com/lists-and-rankings/profit-500/2017-wholesale-distribution-fastest-growing-companies/

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

https://developer.foursquare.com/docs