New Business Recommender System

Nanthakumar Nachimuthu

Jan-2020

1. Introduction

1.1 Background

Starting a new business is big a thing for everyone. It requires good Investment and great Business Plan. I have someone looking for to open a coffee shop in Toronto with questions in mind what kind coffee shop can suite and where to open the shop.

1.2 Problem

The Problem is once decided the business the next ultimate question is 'Where' and 'Which'?. So finding a great location is very important for any business. Using the data we have this project is aim to find a great location also suggest what kind of coffee shop is famous in the city like coffee chains, traditional coffee or local coffee model like that.

1.3 Interest

Any individual who wants to be an entrepreneur will be interested in learning the market trends and potential opportunity for their business.

2. Data Acquisition and Cleaning

2.1 Data Sources

The required data for my project is available on the Web. For Toronto from Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) and I will be using the Foursquare location data for more deep analysis.

The Postal codes from wiki page is like below. It has post code, Borough and Neighborhood info as a tabular format.

Postcode	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Harbourfront
M6A	North York	Lawrence Heights
M6A	North York	Lawrence Manor
M7A	Downtown Toronto	Queen's Park
M8A	Not assigned	Not assigned
M9A	Queen's Park	Not assigned
M1B	Scarborough	Rouge
M1B	Scarborough	Malvern
M2B	Not assigned	Not assigned
M3B	North York	Don Mills North
M4B	East York	Woodbine Gardens
M4B	East York	Parkview Hill

By passing the post codes to Geocodes API to get the Latitude and Longitude values. Then will pass those values into Foursquare API, explore the coffee shops and other business and public gathering areas in each Borough and find out more famous and trendy coffee shops.

2.2 Data Cleaning

1. The first data cleaning is removing the Postal codes if there are no 'Borough's assigned. Also there are few Borough's without Neighborhoods, so assigned the Borough as a Neighborhood in that case.
2. The second level of cleaning is removing the repetitive Venues that queried from Foursquare API. There are high chances of one Venue reported into one or more Neighborhood due to our querying radius. So in this case, I retained the venue records which is closure by Distance to that Neighborhood and removed the other ones.
3. Third cleansing is after we retrieved the Venue Ratings. Found venues without any Ratings, so had to remove those since that may affect our analytics.

2.3 Feature Selection

From Foursquare, I have selected below list of Features for the study.

1. Neighborhood – Name of the Neighborhood
2. Latitude and Longitude of Neighborhood
3. Venue – Name of the Venue
4. Venue ID – Foursquare Venue ID for querying
5. Rating – User rating of the Venue
6. Likes Count – User likes count of the Venue in Foursquare

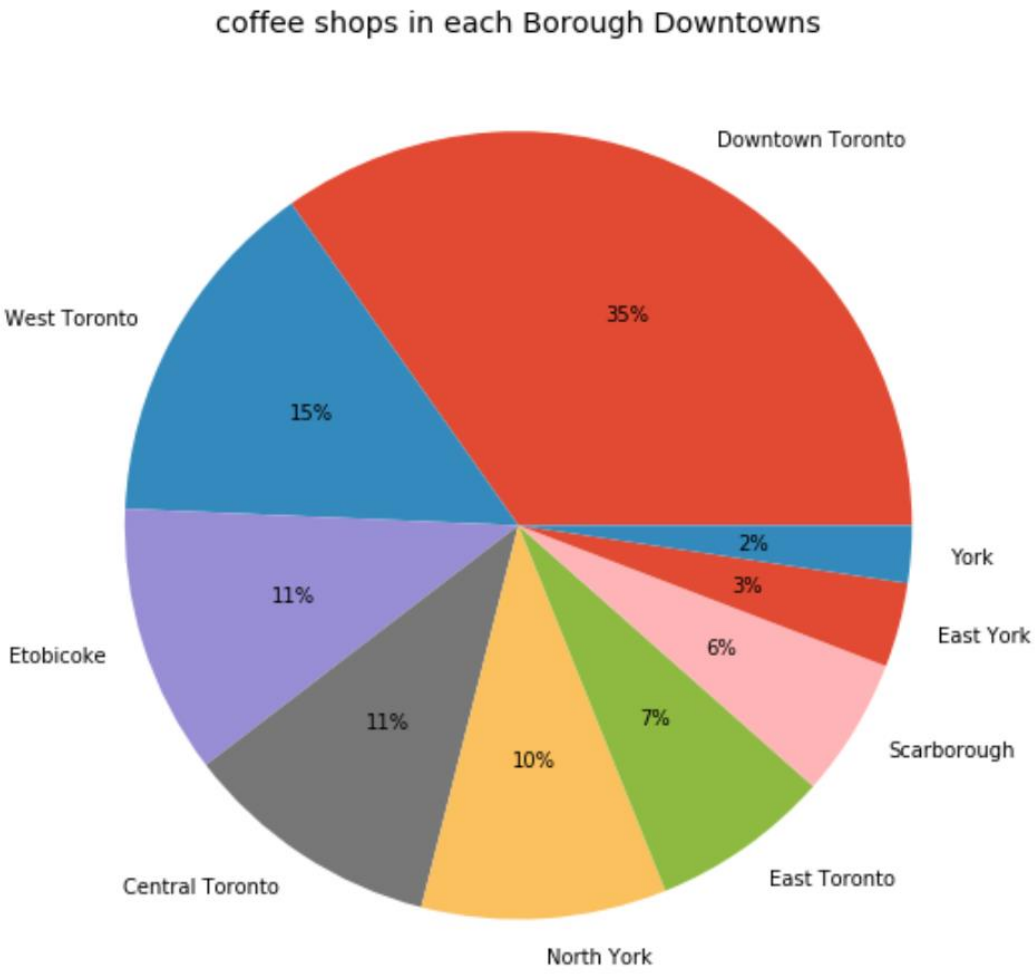
Below are the calculated fields and explanations.

1. Shop Type – At complete Toronto level, using the name of the Venue, if it has 5 or more locations, I classified that as 'Branded' (chain of) coffee shop. And I use the actual name as Shop Type. If fewer than 5 locations found for the Name of the Venue, then I classified as 'Non-Brand'
2. Weight Calculation – I have applyed 30% for Venue Count 50% to Ratings and 20% of weight to Likes for the Venue.

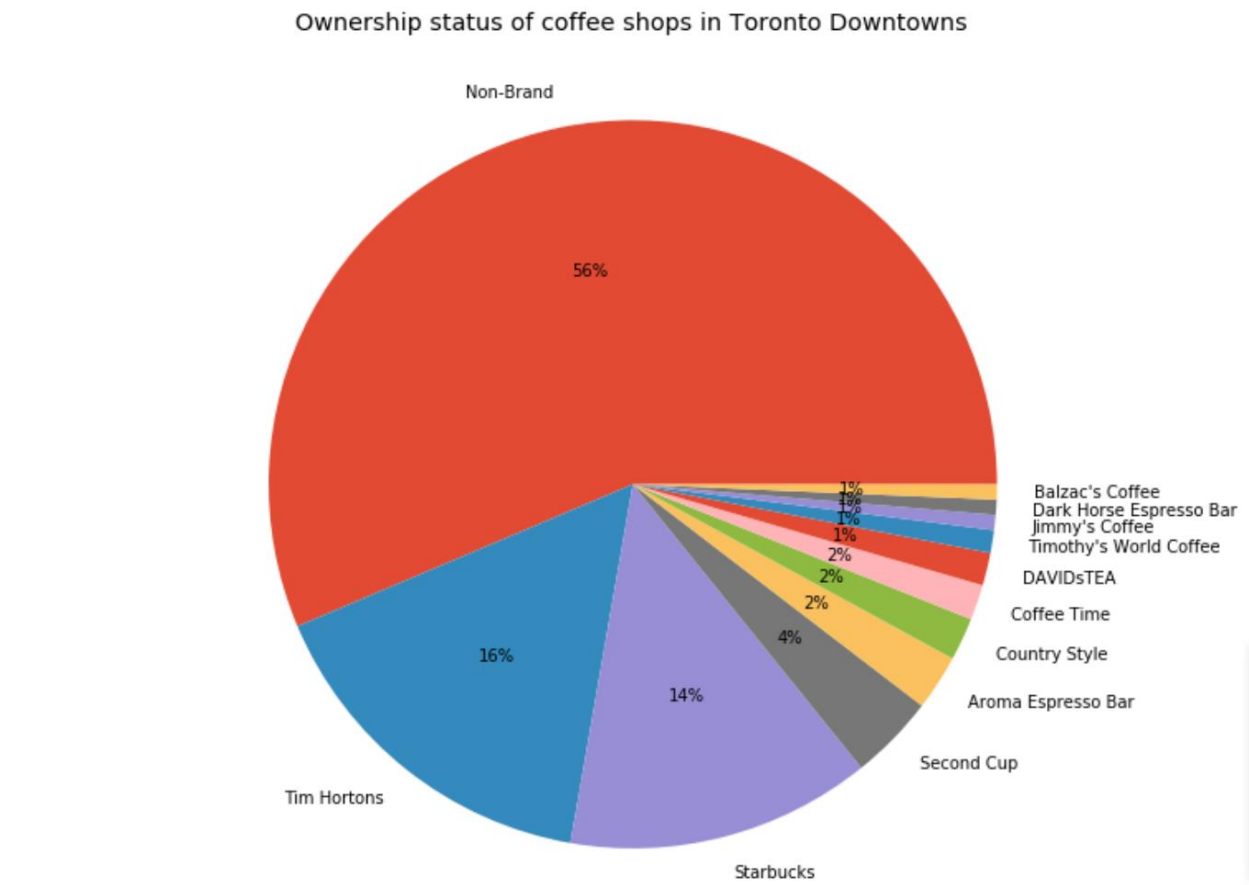
3. Exploratory Data Analysis

Coffee shop density amount the Neighborhoods.

- Obviously Downtown Toronto have 35% of coffee shops out of 894
- West Toronto has 15% and Etobicoke owns 11%
- Below Pie Chat shows the complete picture



Ownership Status of Coffee shops in Toronto



From the above chart, we can say

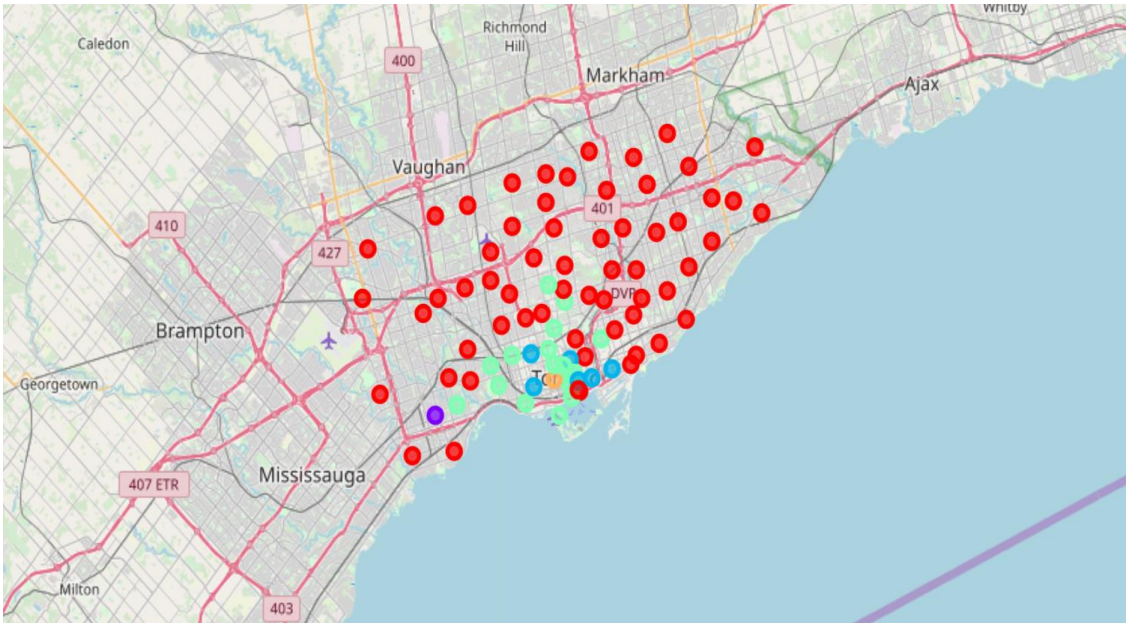
1. **56%** of coffee shops are non-Branded or Local shops
2. **Tim Hortons** owns 16%
3. Starbucks have 14% and Second cup owns around 4 %
4. Rest of 10 % coffee shops are small brands or have 5 to 20 no. of coffee shops in Toronto Area

This is one of great finding – Non-Branded coffee shops are very famous in Toronto.

4. Predictive Modeling

In order to understand more about the Popularity and Users preferences, I have used K-Means Clustering Algorithm. K-Means will segment the coffee shops using the features given and showing the five different clusters.

Illustration of Neighborhoods after segmentation:



Examining Clusters

Over all 'Non-Brand' coffee shops are shows up all the clusters. This means, the Non-Branded coffee shops have mixed ratings and scattered across all over Toronto.

	Cluster Labels	Shop Type	Venue Rating	Venue Count	Neighbourhood
4	0	Tim Hortons	6.350595	48	28
1	0	Non-Brand	7.126212	47	11
3	0	Starbucks	7.071930	30	19
0	0	Aroma Espresso Bar	8.600000	1	1
2	0	Second Cup	7.000000	1	1
5	1	Non-Brand	7.164865	37	1
6	2	Non-Brand	7.949099	84	7
7	3	Non-Brand	7.686889	123	16
8	3	Starbucks	7.348333	22	3
9	4	Non-Brand	7.932258	31	1

5. Conclusion

The below result shows that the 'Non-Branded' coffee shops are rated best by the customers and they are famous in 36 Neighborhood in Toronto area.

Our study recommends, starting a unique coffee shop will give a good success rate than Branded coffee shops. Also, It really doesn't matter the location, since we found the Non-Branded coffee shops are scattered across the Toronto area.

	Shop Type	Venue Rating	Venue Count	Neighbourhood	Cluster Labels
0	Aroma Espresso Bar	8.600000	1	1	1
1	Non-Brand	7.558872	322	36	5
2	Second Cup	7.000000	1	1	1
3	Starbucks	7.109621	52	22	2
4	Tim Hortons	6.350595	48	28	1