OPENING A NEW COFFEE SHOP CHAIN

Capstone Project

Ahmed Lashin

Data Scientist

Company Name Marketing Plan

Heading 3

To get started right away, just tap any placeholder text (such as this) and start typing to replace it with your own.

Want to insert a picture from your files or add a shape, text box, or table? You got it! On the Insert tab of the ribbon, just tap the option you need.

Find even more easy-to-use tools on the Insert tab, such as to add a hyperlink or insert a comment.

HEADING 4

To get started right away, just tap any placeholder text (such as this) and start typing to replace it with your own.

Want to insert a picture from your files or add a shape, text box, or table? You got it! On the Insert tab of the ribbon, just tap the option you need.

Find even more easy-to-use tools on the Insert tab, such as to add a hyperlink or insert a comment.

Introduction

My client is a businessman who is looking to open a new coffee shop chain in the city of Toronto. My client is hoping to build a local chain that can stand out in front of the international coffee chains that are everywhere in Toronto and Canada.

In order to eliminate or at least decrease the competition specially in the early stage of the project, my client would like to open in areas where there are no coffee shops or let us say the minimum competition.

We will conduct a case study to analyze different locations / boroughs in Toronto city to help our customer to select the best location for his new coffee shop chain.

Data

To complete the required analysis, we will need to get data for all districts / boroughs in the city of Toronto, Canada with the geographical coordinates. We will need to get the data of the venues in these boroughs. Specifically, the coffee shops and coffee bars.

The following data will be needed to solve the problem:

- **Neighborhoods in Toronto**
  We will web scrape the Neighborhoods of Toronto using the following Wikipedia page,
  https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M,
  in order to obtain the data that is in the table of postal codes.

- **Latitude and longitude of the neighborhood.**
  We will try to use geopy and geocoder packages to obtain the coordinates data for each neighborhood. Another option is to use the data file that was provided as part of the Capstone from the following link to a csv file that has the geographical coordinates of each postal code:
  http://cocl.us/Geospatial_data

- **Venue data related to coffee shops**
  The data is provided by Foursquare through there APIs.

Methodology

To download the raw dataset, we used read_html from pandas to read in tables from a html, determined the index of raw data and created a data frame.

We then filtered the data to remove all the neighborhoods equal to 'Not Assigned' to get a new dataframe containing only the known neighborhoods.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

We then used the updated dataframe to visualize a map of Toronto with all neighborhoods that will be part of our analysis.

RESTful API Calls to Foursquare

We used the Foursquare API to explore all the districts of the city of Toronto. Although a wide set of venue categories are available, we are interested more in the ones from "Coffee shops" category.

So we get the data of venues for each district and join it with the main district dataframe,

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
4	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant

We then try to aggregate and sort all the venues for each district to get the top 10 venues per categories for each district.
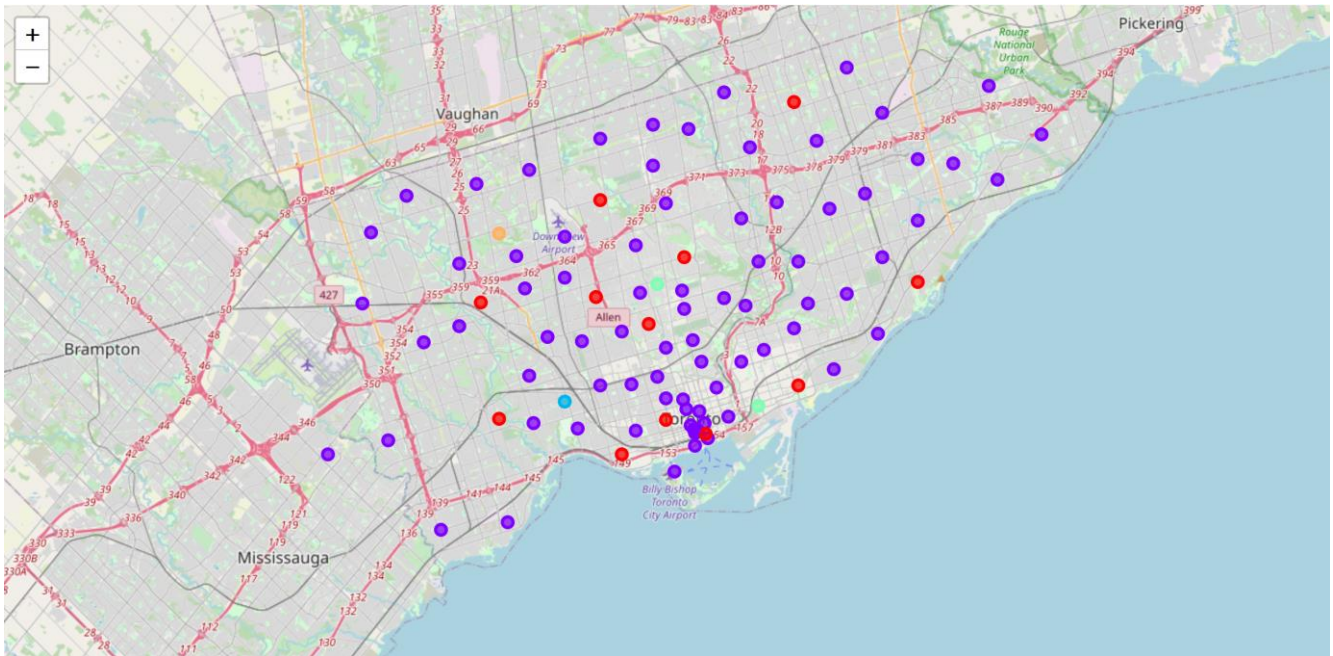
	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Lounge	Skating Rink	Latin American Restaurant	Breakfast Spot	Dumpling Restaurant	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Drugstore
1	Alderwood, Long Branch	Pizza Place	Gym	Pool	Coffee Shop	Skating Rink	Pharmacy	Pub	Sandwich Place	Dessert Shop	Dim Sum Restaurant
2	Bathurst Manor, Wilson Heights, Downsview North	Bank	Coffee Shop	Park	Supermarket	Shopping Mall	Bridal Shop	Sandwich Place	Restaurant	Pizza Place	Pharmacy
3	Bayview Village	Café	Japanese Restaurant	Bank	Chinese Restaurant	Yoga Studio	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop
4	Bedford Park, Lawrence Manor East	Coffee Shop	Sandwich Place	Italian Restaurant	Grocery Store	Sushi Restaurant	Greek Restaurant	Pharmacy	Pizza Place	Pub	Café

We then try to apply K-Means algorithm in order to cluster the districts and conduct more analysis.
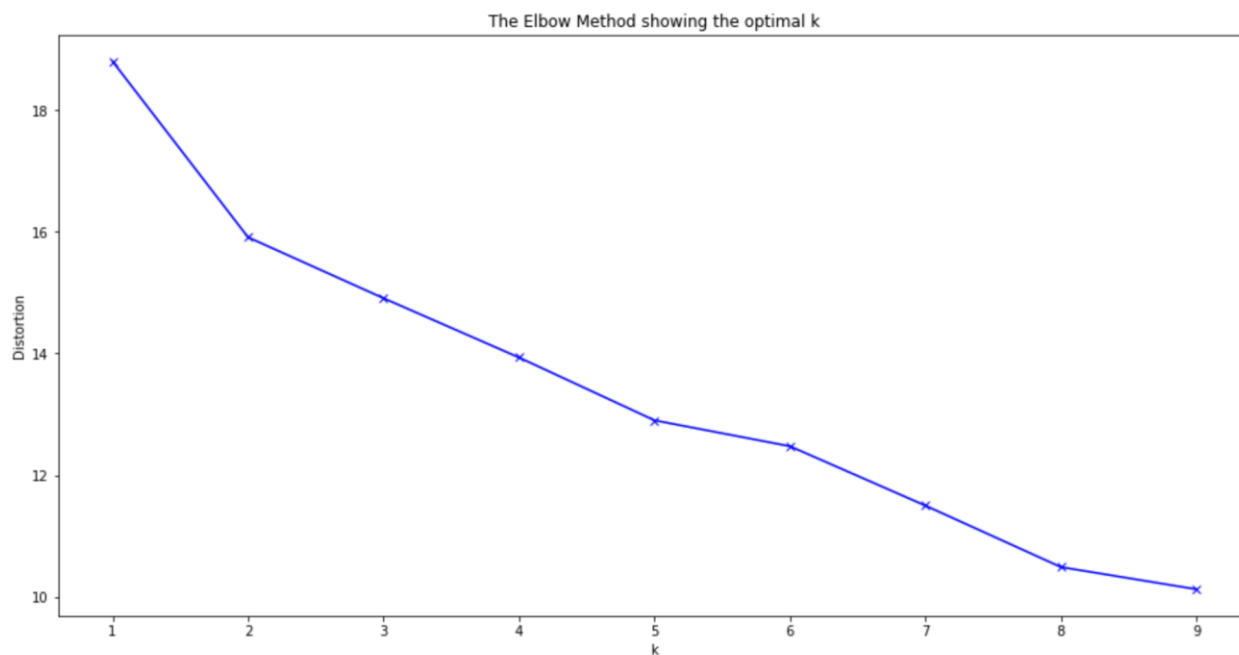
We found that the optimal number of clusters is 5.

We then apply the clustering algorithm on the data to cluster all the districts into 5 clusters. We then visualize a map of Toronto for all districts as per their clusters.

Results

Cluster 0

As we can see in the results for cluster 0, it contains a lot of districts and many venues. Coffee shops and restaurants are among the top 10 venues.

	Borough	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	North York	0	Park	Food & Drink Shop	Yoga Studio	Drugstore	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Eastern European Restaurant
21	York	0	Park	Women's Store	Pool	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Department Store
35	East York	0	Park	Convenience Store	Intersection	Yoga Studio	Donut Shop	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant
52	North York	0	Park	Yoga Studio	Drugstore	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant
61	Central Toronto	0	Park	Bus Line	Swim School	Doner Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Donut Shop	Farmers Market
64	York	0	Park	Convenience Store	Yoga Studio	Drugstore	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Eastern European Restaurant
66	North York	0	Park	Convenience Store	Yoga Studio	Drugstore	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Eastern European Restaurant
77	Etobicoke	0	Park	Mobile Phone Shop	Sandwich Place	Doner Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Donut Shop	College Stadium
85	Scarborough	0	Park	Coffee Shop	Playground	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant
91	Downtown Toronto	0	Park	Playground	Trail	Doner Restaurant	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Donut Shop

Cluster 1

As we can see in the results for cluster 1, it contains a lot of districts and many venues. Coffee shops and restaurants are among the top 10 venues.

	Borough	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	North York	1	Intersection	Pizza Place	Coffee Shop	Hockey Arena	Portuguese Restaurant	Doner Restaurant	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store
2	Downtown Toronto	1	Coffee Shop	Park	Bakery	Café	Pub	Breakfast Spot	Restaurant	Theater	Spa	Shoe Sto
3	North York	1	Clothing Store	Accessories Store	Event Space	Boutique	Furniture / Home Store	Coffee Shop	Vietnamese Restaurant	Donut Shop	Discount Store	Distributi Center
4	Downtown Toronto	1	Coffee Shop	Yoga Studio	Sushi Restaurant	Bar	Beer Bar	Smoothie Shop	Sandwich Place	Café	Portuguese Restaurant	Chinese Restaura
6	Scarborough	1	Fast Food Restaurant	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Yoga Studio	Dessert S
7	North York	1	Gym	Restaurant	Coffee Shop	Japanese Restaurant	Beer Store	Italian Restaurant	Sporting Goods Shop	Bike Shop	Sandwich Place	Café
13	North York	1	Gym	Restaurant	Coffee Shop	Japanese Restaurant	Beer Store	Italian Restaurant	Sporting Goods Shop	Bike Shop	Sandwich Place	Café
8	East York	1	Pizza Place	Gym / Fitness Center	Café	Athletics & Sports	Gastropub	Bank	Intersection	Pharmacy	Pet Store	Breakfas Spot
9	Downtown Toronto	1	Coffee Shop	Clothing Store	Café	Bubble Tea Shop	Cosmetics Shop	Japanese Restaurant	Middle Eastern Restaurant	Lingerie Store	Italian Restaurant	Pizza Pla
10	North York	1	Sushi Restaurant	Japanese Restaurant	Pizza Place	Asian Restaurant	Pub	Yoga Studio	Dog Run	Diner	Discount Store	Distributi Center
12	Scarborough	1	Bar	Yoga Studio	Drugstore	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant	Field

Cluster 2

As we can see in the results for cluster 2, it contains only 1 district. There are no Coffee shops among the top 10 venues.

	Borough	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
62	Central Toronto	2	Garden	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Yoga Studio	Fast Food Restaurant

Cluster 3

As we can see in the results for cluster 3, it contains only 2 district. There are no Coffee shops among the top 10 venues.

	Borough	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
32	Scarborough	3	Playground	Yoga Studio	Drugstore	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant
83	Central Toronto	3	Park	Lawyer	Playground	Donut Shop	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Dumpling Restaurant

Cluster 4

As we can see in the results for cluster 4, it contains only 1 district. There are no Coffee shops among the top 10 venues.

	Borough	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
57	North York	4	Baseball Field	Yoga Studio	Drugstore	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant	Field

Conclusions

The analysis of coffee shop venues around the Toronto during this project showed that some districts are highly active and contains too manty venues like cluster 0 and 1. The coffee shops are among the top 10 venues in the districts of these clusters.

While other clusters like 2, 3 and 4 are less active. In fact, some of these districts does not even have coffee shops among the top 10 categories of venues.

So if the goal of our customer is to stay away from the competition, He should utilize the information in this analysis to decide which districts should be used to start the new coffee shop chain to minimize the competition.

All districts in clusters 2, 3 and 4 can be a good location to start.

A complementary analysis to further explore this subject in the future could include the consideration of the population and demographic distribution for the people living in these districts in order to give more clear idea about the potential customers for the new coffee shop chain and their spending habits.

References

A Jupyter Notebook with complete code for this analysis can be found in the following link,

https://github.com/asLashin/Coursera_Capstone/blob/master/Capstone%20Project:%20opening%20a%20New%20CoffeeShop%20Chain%20C.ipynb