Coursera Capstone Project

The Battle of Neighbourhoods

Introduction

Business Problem

Toronto is the capital city of the Canadian province of Ontario and is the most populous city in Canada with a figure of approximately 2.93 million. This figures constitutes a well-diversified city leading to a plethora of business opportunities in a variety of markets.

With a well-developed city, such as Toronto, this means that markets will be highly competitive causing the need for analysis in order to identify if there is a gap in the market for a particular business idea. With thorough analysis, we can highlight how saturated the market is and also gauge an idea of the risk involved.

Problem Description

An Indian Restaurant (IndyRest) is looking to bring itself to the market in the city of Toronto and is in need of in depth analysis to decide on where is best to locate their first restaurant in order to build its brand and reputation with the view to build a chain of restaurants across Canada.

Key factors

To determine the best location or IndyRest, we will examine the demand for such a business idea through key factors such as the neighbourhood density of the surrounding boroughs. Another key factor is to analyse the competitiveness of the market for restaurants and, in particular, Indian restaurants so we can highlight the market saturation levels.

Audience

This analysis is targeted towards the management team of IndyRest to assist them in choosing a prime location for the restaurant and also to give them other choices with a range of locations and attributes associated with these locations which they may feel to be beneficial financially for their restaurant.

Data description

Data URL - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

We will use the Scarborough dataset that was scraped from Wikipedia on Week 3.
This dataset should provide us with details such as latitude and longitude values and zip codes which is paramount to helping us pick the right location.

We will need to use Foursquare API locational information in order to gain data regarding different venues in the neighbourhoods of each borough. Foursquare API will also provide us with various categories associated with that neighbourhood such as common venues which is a key factor in picking the best location for high demand and lowest competition.

The information gathered from each neighbourhood will be as follows:

1. Neighbourhood
2. Neighbourhood Latitude
3. Neighbourhood Longitude
4. Venue Name
5. Venue Category
6. Venue Latitude
7. Venue Longitude

How will this data be used

With these features, along with the techniques such as K-clustering to segment and cluster the neighbourhoods, we will be able to provide IndyRest with the best possible neighbourhood location to start off the beginning of their journey. By analysing the most common venues and their categories in each neighbourhood, this will give us an idea of the competition levels as well as if there is a gap in the market for this particular venture. We will want to find a borough which is neighbourhood dense and preferably with the least amount of restaurants in the vicinity which lowers competition levels making the restaurant a market leader in this location. Other additional features which could improve demand and add value is to analyse proximity to the coast which could provide a key selling point for the restaurant.

Methodology

We are required to find the best location for IndyRest to start their new restaurant in Toronto city.

Firstly, we needed three types of geolocational data. The first is postal codes in the Toronto area in order to combine boroughs with neighbourhoods. The second is to obtain the coordinates of these neighbourhoods and the third is to apply this to the data of restaurants in Toronto. To achieve this, data scrapping the Wikipedia page for postal codes was necessary. I then clean the data, getting rid of Not assigned values and out of city Toronto boroughs in order to create data ready for informative analysis.

Also, we need coordinates for these neighbourhoods so we used the geocoder API to obtain this data. This was necessary for the geolocational coordinates of each borough so we can merge with the first data frame as well as applying these coordinates for each restaurant in Toronto city. The geocoder API was essential in enabling us to input information into Foursquare.com to provide us with venue information in each neighbourhood.

We will also use techniques such as Kmeans clustering to segment and cluster Toronto neighbourhoods in order to allow us to distinguish between which neighbourhoods are densely populated as the higher the frequency of neighbourhoods the higher the likelihood of demand which is important for a new restaurant coming to the market and trying to create awareness.

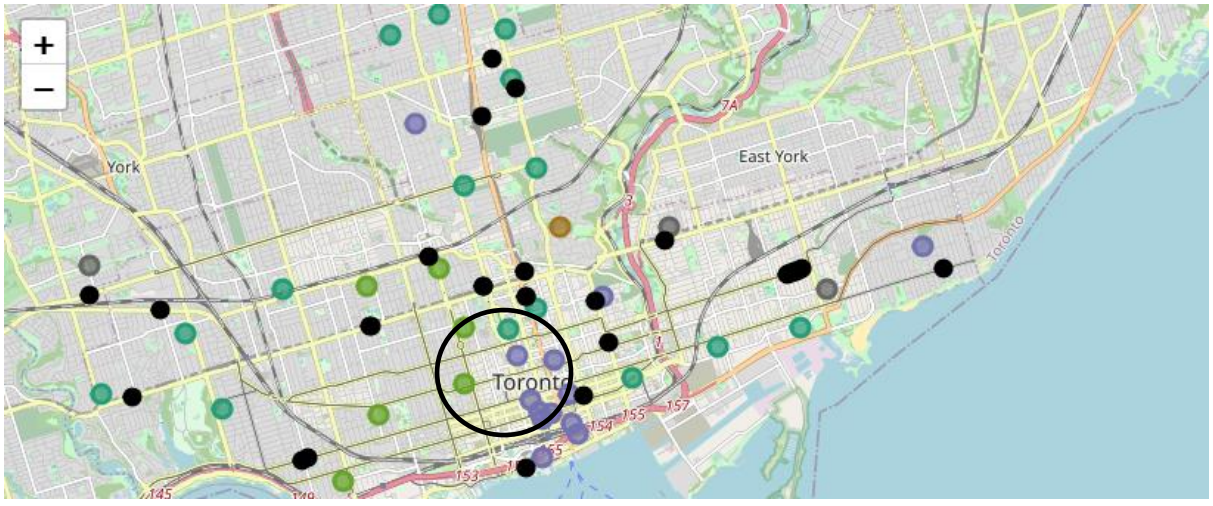
We would also use the Foursquare API to identify the frequency of restaurant cuisines in each neighbourhood as well as the top 10 common food cuisines in each neighbourhood with the use of Kmeans clustering. The usage of Foursquare API also allowed me to identify the areas in which Indian restaurants were already situated.

Visualisation will also be provided using the folium package to map out the locations of current Indian restaurants so we can locate further away from them to reduce competition. We also used visualisation to map out clusters of neighbourhoods in Toronto to highlight areas of Toronto which are neighbourhood dense.

With these methodologies, we will be able to provide IndyRest with the best recommendation for their new restaurant which will be a location consisting of low direct competition levels and neighbourhood dense areas.

Results

With Kmeans clustering, we was able to cluster the neighbourhoods based on their five boroughs. Through the visualisation, we are able to observe that the black markers, which represent Downtown Toronto, are neighbourhood dense and so consequently we will focus on locating the restaurant around Downtown Toronto for demand reasons.



Through this visualisation it is clear to see that the restaurant market in the city of Toronto is very competitive with restaurants scattered across the city in various locations. The black markers represent areas which already contain a Indian restaurant. The visualisation highlights that Indian restaurants are distributed all across the city of Toronto and therefore competition in this specific market is very fierce and therefore choosing a location will prove very difficult. Depending on the marketing strategy of IndyRest, they may choose to enter

where there is already competition in order to get rid of its competitors and be the market leader however this would be very difficult for a start-up. From the mapping, I would recommend that IndyRest would locate in Downtown Toronto around the area circled on the map which gives them the least amount of direct competition as well as the most demand in such a neighbourhood dense area.

	Borough	Neighbourhood	Latitude	Longitude	Cluster_ID	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
2	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	2.0	Japanese Restaurant	Restaurant	Italian Restaurant	Seafood Restaurant	Ramen Restaurant	Middle Eastern Restaurant	German Restaurant	Vegetarian / Vegan Restaurant	Thai Restaurant	Sushi Restaurant
3	Downtown Toronto	St. James Town	43.651494	-79.375418	2.0	Restaurant	Italian Restaurant	Seafood Restaurant	American Restaurant	Japanese Restaurant	Latin American Restaurant	Thai Restaurant	New American Restaurant	Asian Restaurant	Middle Eastern Restaurant
4	East Toronto	The Beaches	43.676357	-79.293031	2.0	Japanese Restaurant	Asian Restaurant	Caribbean Restaurant	Restaurant	Mexican Restaurant	Indian Restaurant	Greek Restaurant	French Restaurant	Ramen Restaurant	Mediterranean Restaurant
5	Downtown Toronto	Berczy Park	43.644771	-79.373306	2.0	Japanese Restaurant	Restaurant	Seafood Restaurant	French Restaurant	Vegetarian / Vegan Restaurant	Thai Restaurant	New American Restaurant	Middle Eastern Restaurant	Italian Restaurant	American Restaurant
6	Downtown Toronto	Central Bay Street	43.657952	-79.387383	2.0	Japanese Restaurant	Ramen Restaurant	Vegetarian / Vegan Restaurant	Sushi Restaurant	Seafood Restaurant	Mexican Restaurant	Falafel Restaurant	Thai Restaurant	Restaurant	Modern European Restaurant
8	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568	2.0	Japanese Restaurant	Sushi Restaurant	Fast Food Restaurant	Seafood Restaurant	Restaurant	Italian Restaurant	American Restaurant	Mediterranean Restaurant	Brazilian Restaurant	Vegetarian / Vegan Restaurant

With the use of Foursquare.com API, we are able to leverage on the data to analyse neighbourhoods in the Downtown Toronto area and to find out their top ten common restaurants. This is essential for us to pick a neighbourhood with no Indian restaurant in the top ten in order to pick a location with the least direct competition.

Discussion

Based on the two visualisations, I recommend that IndyRest should locate anywhere in the circle, ideally closer to the centre of the circle which makes sure they are a good distance away from direct competition. Through analysing the circle created, it is clear that the purple clusters are most frequent which led me to choose a neighbourhood in that specific cluster. I had narrowed the choice of neighbourhoods down to two instead of one in order to give IndyRest the freedom of choice instead of just one option. These neighbourhoods are Central Bay street and Richmond Adelaide, King. Personally, I would tend towards Richmond Adelaide, King due to it being in close proximity towards the coast which is a source of added value to the restaurant.
From the top ten most common venues, Indian Restaurant does not fall into this list which is very promising for IndyRest.

Conclusion

To conclude, the best recommendation for IndyRest to locate in Toronto will be the neighbourhood of Richmond Adelaide, King. This neighbourhood encapsulated all the important factors that we were looking in terms of being neighbourhood dense for demand reasons and also not having many Indian Restaurants in the proximity thus decreasing competition levels.
We also provided IndyRest with a second option which is up to their discretion and therefore providing them with options so they don't feel obligated to pick this location if it is not to their preference.