Predicting neighbourhood for starting Indian Restaurant

1. Introduction

1.1 Background

Toronto is one of the prominent places to start any new venture. Given that there is rise in immigrants and tourists, we present a report to stakeholders holding analysis of the various cuisines available in the city center of Toronto and understand the area available to submit the

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an Indian restaurant in Toronto, Canada.
Since there are lots of restaurants in Toronto, we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no Indian restaurants in vicinity. We would also prefer locations as close to city center as possible, if first two conditions are met. we are using Nathan Plaza Square as our city center of Toronto
We will use our data science powers to generate a few most promising neighbourhoods based on these criteria.
Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

2. Data acquisition and Cleansing

2.1 Sources

For the analysis of this project, we have scraped the data from Wikipedia . Although, this data has the postal codes in Canada, we don't have the coordinates for these boroughs.

2.2 Data Cleansing

The data available in Wikipedia page had many neighbourhoods as "not assigned" against each borough. Besides, the value of borough against certain postal code had values "Not assigned". We first removed the rows which had the boroughs as "not assigned" and assigned value of the borough as that of the value of neighbourhood whose values were "not assigned". Later, the neighbourhoods were grouped based on the postal code and assigned to a dataframe. The coordinates were assigned to another dataframe and these two dataframes were merged based on the postcode.

The dataframe containing both the post codes and coordinates were filtered with the keyword "Toronto", and the resultant dataframe was cast onto a map using folium to visualize the demographics and the frequency of venues in each neighbourhood around Toronto. As "Nathan Plaza Square" is located in the center of Toronto, we have used this location as the hub and started exploring neighbourhoods of radius ~2.Kms from city center to find if there are any neighbourhoods that has more number of tourists coming in, closest to the city center, and there are no Indian restaurants in the vicinity.

Based on the above criteria, **Kensington Market and Cabbage Town** came top in the list. Later, these areas were explored using folium to arrive at the conclusion that these two neighbourhoods are best suited to start an Indian restaurant, as these neighbourhoods are revered diversified in nature, and people around the globe come to visit the Victorian homes and explore the different cuisines available at their best.

3. Exploratory Data Analysis

Based on definition of our problem, factors that will influence our decision are:

- Number of existing restaurants in the neighbourhood (any type of restaurant)
- Number of and distance to Indian restaurants in the neighbourhood, if any
- Distance of neighbourhood from city center we decided to use regularly spaced grid of locations, centered around city center, to define our neighbourhoods.

Following data sources will be needed to extract/generate the required information:

- Number of restaurants and their type and location in every neighbourhood will be obtained using Foursquare API. Coordinates of Toronto center will be obtained using Google Maps API geocoding of well-known Toronto location (Nathan Plaza Square)
- Neighbourhood Candidates
  We create latitude & longitude coordinates for centroids of our candidate neighbourhoods. We will create a grid of cells covering our area of interest which is approx. 12x12 kilometres centered around Toronto city center
- Distance of neighbourhoods from Center

We then created a grid of area candidates, equally spaced, centered around city center and within ~6km from Nathan Plaza Square. Our neighbourhoods will be defined as circular areas with a radius of 300 meters, so our neighbourhood centers will be 600 meters apart.

3.1. 2D representation to calculate distance

To accurately calculate distances, we need to create our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on Folium map. So, we will create functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).
We then create a hexagonal grid of cells: we offset every other row, and adjust vertical row spacing so that every cell center is equally distant from all its neighbours.

We now have the coordinates of centers of neighbourhoods/areas to be evaluated, equally spaced (distance from every point to its neighbours is the same) and within ~6km from Nathan Plaza Square.

We then calculate the neighbourhoods and their distance from center as depicted below.

	Address	Latitude	Longitude	X	Y	Distance from center
0	490 Unwin Ave, ON M5A, Canada	43.645435	-79.332848	-5.312457e+06	1.050178e+07	5992.495307
1	440 R Unwin Ave, ON M4M 3B9, Canada	43.649160	-79.333409	-5.311857e+06	1.050178e+07	5840.376700
2	515 Commissioners St, ON M4M 1A5, Canada	43.652885	-79.333971	-5.311257e+06	1.050178e+07	5747.173218
3	10 Lower Don Recreation Trail, ON M4M 1E4, Canada	43.656611	-79.334532	-5.310657e+06	1.050178e+07	5715.767665
4	37 Winnifred Ave, ON M4M 2X2, Canada	43.660337	-79.335093	-5.310057e+06	1.050178e+07	5747.173218
5	65 Bertmount Ave, ON M4M 2X8, Canada	43.664063	-79.335655	-5.309457e+06	1.050178e+07	5840.376700
6	46 Galt Ave, ON M4M 2Z1, Canada	43.667790	-79.336216	-5.308857e+06	1.050178e+07	5992.495307
7	4 Regatta Rd, ON M5A, Canada	43.639494	-79.336449	-5.313357e+06	1.050230e+07	5855.766389
8	315 Unwin Ave, ON M4M 3M9, Canada	43.643219	-79.337010	-5.312757e+06	1.050230e+07	5604.462508
9	17 Basin St, ON M4M 1A1, Canada	43.646944	-79.337572	-5.312157e+06	1.050230e+07	5408.326913

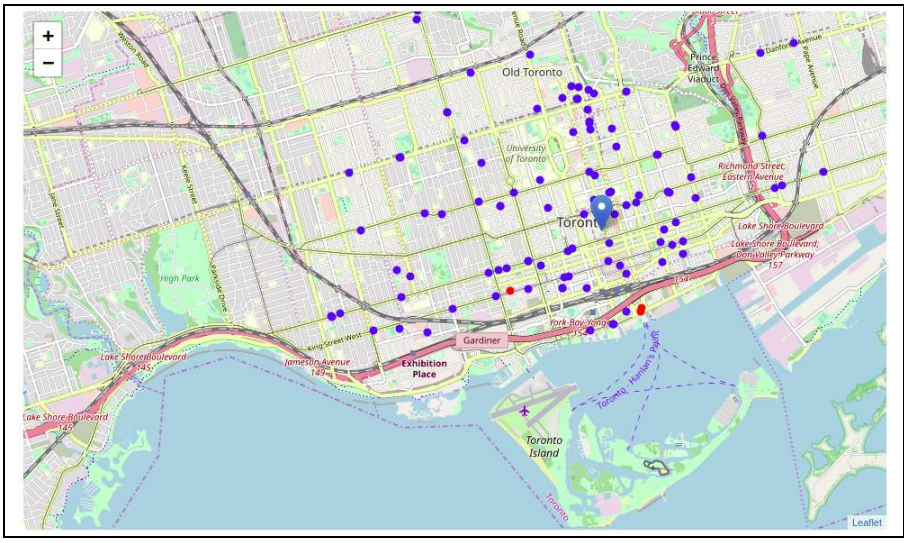
3.2 Using Foursquare API to get the venues

Now that we have our location candidates, let's use Foursquare API to get info on restaurants in each neighbourhood. We're interested in venues in 'food' category, but only those that are proper restaurants - coffee shops, pizza places, bakeries etc. are not direct competitors so we don't care about those. So, we will include in our list only venues that have

'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific 'Italian restaurant' category, as we need info on Italian restaurants in the neighbourhood.
Using the category ID available in foursquare portal for Indian restaurants, we will use it find the Indian restaurants available out of the total restaurants.

Total number of restaurants	107
Total number of Indian restaurants	3
Percentage of Indian restaurants	2.80%
Average number of restaurants in neighbourhood	0.475

Let's now see all the collected restaurants in our area of interest on map, and let's also show Indian restaurants in different colour.



Now we have all the restaurants in area within few kilometres from Nathan Plaza square, and we know which ones are Indian restaurants! We also know which restaurants exactly are in vicinity of every neighbourhood candidate center.
This concludes the data gathering phase - we're now ready to use this data for analysis to produce the report on optimal locations for a new Indian restaurant!

4. Methodology

In this project we will direct our efforts on detecting areas of Toronto that have low restaurant density, particularly those with low number of Indian restaurants. We will limit our analysis to area ~6km around city center.
In first step we have collected the required data: location and type (category) of every restaurant within 6km from Toronto center (Nathan Square Plaza). We have also identified Indian restaurants (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of Berlin - we will use heatmaps to identify a few promising areas close to center with low number of restaurants in general (and no Italian restaurants in vicinity) and focus our attention on those areas.
In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants in radius of 250 meters, and we want locations without Indian restaurants in radius of 400 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighbourhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

5. Analysis

Let's perform some basic explanatory data analysis and derive some additional info from our raw data. First let's count the number of restaurants in every area candidate:

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area
0	490 Unwin Ave, ON M5A, Canada	43.645435	-79.332848	-5.312457e+06	1.050178e+07	5992.495307	0
1	440 R Unwin Ave, ON M4M 3B9, Canada	43.649160	-79.333409	-5.311857e+06	1.050178e+07	5840.376700	0
2	515 Commissioners St, ON M4M 1A5, Canada	43.652885	-79.333971	-5.311257e+06	1.050178e+07	5747.173218	0
3	10 Lower Don Recreation Trail, ON M4M 1E4, Canada	43.656611	-79.334532	-5.310657e+06	1.050178e+07	5715.767665	0
4	37 Winnifred Ave, ON M4M 2X2, Canada	43.660337	-79.335093	-5.310057e+06	1.050178e+07	5747.173218	0
5	65 Bertmount Ave, ON M4M 2X8, Canada	43.664063	-79.335655	-5.309457e+06	1.050178e+07	5840.376700	0
6	46 Galt Ave, ON M4M 2Z1, Canada	43.667790	-79.336216	-5.308857e+06	1.050178e+07	5992.495307	0
7	4 Regatta Rd, ON M5A, Canada	43.639494	-79.336449	-5.313357e+06	1.050230e+07	5855.766389	0
8	315 Unwin Ave, ON M4M 3M9, Canada	43.643219	-79.337010	-5.312757e+06	1.050230e+07	5604.462508	0
9	17 Basin St, ON M4M 1A1, Canada	43.646944	-79.337572	-5.312157e+06	1.050230e+07	5408.326913	0

OK, now let's calculate the distance to nearest Indian restaurant from every area candidate center (not only those within 300m - we want distance to closest one, regardless of how distant it is).

	Address	Latitude	Longitude	X	Y	Distance from center	Restaurants in area	Distance to Indian restaurant
0	490 Unwin Ave, ON M5A, Canada	43.645435	-79.332848	-5.312457e+06	1.050178e+07	5992.495307	0	5020.761297
1	440 R Unwin Ave, ON M4M 3B9, Canada	43.649160	-79.333409	-5.311857e+06	1.050178e+07	5840.376700	0	5053.976202
2	515 Commissioners St, ON M4M 1A5, Canada	43.652885	-79.333971	-5.311257e+06	1.050178e+07	5747.173218	0	5157.257691
3	10 Lower Don Recreation Trail, ON M4M 1E4, Canada	43.656611	-79.334532	-5.310657e+06	1.050178e+07	5715.767665	0	5326.531549
4	37 Winnifred Ave, ON M4M 2X2, Canada	43.660337	-79.335093	-5.310057e+06	1.050178e+07	5747.173218	0	5555.769055
5	65 Bertmount Ave, ON M4M 2X8, Canada	43.664063	-79.335655	-5.309457e+06	1.050178e+07	5840.376700	0	5837.910691
6	46 Galt Ave, ON M4M 2Z1, Canada	43.667790	-79.336216	-5.308857e+06	1.050178e+07	5992.495307	0	6165.698069
7	4 Regatta Rd, ON M5A, Canada	43.639494	-79.336449	-5.313357e+06	1.050230e+07	5855.766389	0	4591.028752
8	315 Unwin Ave, ON M4M 3M9, Canada	43.643219	-79.337010	-5.312757e+06	1.050230e+07	5604.462508	0	4512.543223
9	17 Basin St, ON M4M 1A1, Canada	43.646944	-79.337572	-5.312157e+06	1.050230e+07	5408.326913	0	4509.731453

Average distance to closest Indian restaurant from each area center: 3517.7153824688994

OK, so on average Indian restaurant can be found within ~5500m from every area center candidate.
Let's create a map showing heatmap / density of restaurants and try to extract some meaningful info from that. Also, let's show borders of Toronto boroughs on our map and a few circles indicating distance of 1km, 2km and 3km from Nathan Plaza Square.

Below depiction is the total number of restaurants around city center, and the total number of Indian restaurants in Toronto

We see low restaurant density closest to city center on North-west and North-east.

Indian restaurants represent a subset of ~2% of all restaurants in Toronto, but it also indicates existing Indian restaurants in South and South-west from Nathan Plaza Square, with closest pockets of low Indian restaurant density positioned North, North-west from city center.
Based on this we will now focus our analysis on areas North, North-west from Toronto center - we will move the center of our area of interest and reduce its size to have a radius of 2.5km. This places our location candidates mostly in boroughs **Kensington Market** and **Cabbage town**

5.1. Kensington Market and Cabbage town

Analysis of popular travel guides and web sites often mention Kensington Market is a cheerful bustle on summer's days, beautiful, interesting, rich with culture, cool neighbourhood popular among tourists. Both Kensington Market and Cabbage town are one of Toronto's oldest and most diverse neighbourhoods.

"Kensington Market has long been revered for its diverse cultural life and as a part of Toronto. With bars, coffee shops, restaurants, Organic grocery stores, bookstores, vintage clothing stores and all manner of other little oddities. The market is as eclectic as the people who live there, the neighbourhood is varied, but the music is constant. Each restaurant produces a smell from a different corner of the world, fusing flavours that have never been put together before. The peace and the energy are frenetic. Kensington market has also been a favourite spot for tourists as the locality has cuisines from around the globe." (Airbnb)

"Cabbage town is one of the few neighbourhoods in Toronto that has really kept the essence of its spirit as the world around it has changed. Most of the stores and restaurants are mom and pop stores, giving the whole area a quiet, community feel. Nothing proves the point more than the working farm you can find in the middle of Cabbage town's mix of Victorian homes and global cuisine. It's located in Toronto's East End, so both the Beaches and Downtown Toronto are a short stroll away." (theculturetrip.com)

	Latitude	Longitude	X	Y	Restaurants nearby	Distance to Indian restaurant
0	43.657556	-79.358182	-5.310207e+06	1.050449e+07	0	3206.338303
1	43.658177	-79.358276	-5.310107e+06	1.050449e+07	0	3276.641149
2	43.654081	-79.358405	-5.310757e+06	1.050458e+07	0	2782.019216
3	43.654702	-79.358499	-5.310657e+06	1.050458e+07	0	2843.484276
4	43.655323	-79.358593	-5.310557e+06	1.050458e+07	0	2907.090424
5	43.655944	-79.358687	-5.310457e+06	1.050458e+07	0	2972.700228
6	43.656565	-79.358781	-5.310357e+06	1.050458e+07	0	3040.183967
7	43.657186	-79.358876	-5.310257e+06	1.050458e+07	0	3109.419634
8	43.657807	-79.358970	-5.310157e+06	1.050458e+07	0	3180.292811
9	43.658428	-79.359064	-5.310057e+06	1.050458e+07	0	3252.696462

OK. We will now filter those locations: we're interested only in locations with no more than two restaurants in radius of 250 meters, and no Indian restaurants in radius of 400 meters. Below are the findings.

Locations with no more than two restaurants nearby	2197
Locations with no Indian restaurants within 400m	2214
Locations with both conditions met	2156

We now have a bunch of locations fairly close to Nathan Plaza Square (mostly in Kensington, Cabbage Town), and we know that each of those locations has no more than two restaurants in radius of 250m, and no Indian restaurant closer than 4000m. Any of those locations is a potential candidate for a new Indian restaurant, at least based on nearby competition. Looking at the below heatmaps, we have now is a clear indication of zones with low number of restaurants in vicinity, and no Indian restaurants at all nearby.

6. Clustering the neighbourhood

We now cluster those locations to create centers of zones containing good locations. Those zones, their centers and addresses will be the final result of our analysis.

Our clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.

```
11 King St W, Toronto, ON M5H 4C7                  => 0.9km from Nathan Plaza Square
444 Yonge St, Toronto, ON M5B 2H4                  => 1.3km from Nathan Plaza Square
11 St Bartholomew St, Toronto, ON M5A 1Y5          => 2.5km from Nathan Plaza Square
85 Huron St, Toronto, ON M5T 2A8                   => 1.5km from Nathan Plaza Square
Parliament St at Winchester St, Toronto, ON M4X 1P4 => 2.7km from Nathan Plaza Square
250 Front St E, Toronto, ON M5A 1E9                => 2.3km from Nathan Plaza Square
Teefy Hall, Toronto, ON M5S 2C4                    => 2.3km from Nathan Plaza Square
561 Jarvis St, Toronto, ON M4Y 3B9                 => 2.5km from Nathan Plaza Square
161 Gerrard St E, Toronto, ON M5A 2E4              => 1.7km from Nathan Plaza Square
Union Station, 65 Front St W, Toronto, ON M5J 1E6  => 1.4km from Nathan Plaza Square
21 Lower Jarvis St, Toronto, ON M5E 1R8            => 1.9km from Nathan Plaza Square
136 Jarvis St, Toronto, ON M5B 2B5                 => 1.1km from Nathan Plaza Square
Sandford Fleming Bldg, Toronto, ON M5S 3G4         => 1.8km from Nathan Plaza Square
46 Widmer St, Toronto, ON M5V 2E9                  => 1.2km from Nathan Plaza Square
89 Chestnut Street Residence, 89 Chestnut St, Toronto, ON M5G 1R1 => 0.3km from Nathan Plaza Square
```

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of restaurants and no Italian restaurants nearby, all zones being fairly close to city center (all less than 3km from Nathan Plaza Square, and about half of those less than 2km from city center). Although zones are shown on map with a radius of ~5000 meters (green circles), their shape is actually very irregular, and their centers/addresses should be considered only as a starting point for exploring area neighbourhoods in search for potential restaurant locations. Most of the zones are located in Kensington Market and Cabbage Town boroughs, which we have identified as interesting due to being popular with tourists, fairly close to city center and well connected by public transport.

7. Results and Discussion

Our analysis shows that although there are quite a number of restaurants in Toronto (~350 in our initial area of interest which was 12x12km around Nathan Plaza Square), there are pockets of low restaurant density fairly close to city center. Highest concentration of restaurants was detected South and South-west from Nathan Plaza Square, so we focused our attention to areas south, south-east and east, corresponding to boroughs Kensington Market and Cabbage Town.
After directing our attention to this narrower area of interest (covering approx. 5x5km North-west from Nathan Plaza Square) we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with an Italian restaurant closer than 400m were removed.
Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.
Result of all this is 15 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Indian restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Toronto center but not crowded with existing restaurants (particularly Indian) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition, but also other factors considered, and all other relevant conditions met.

8. Conclusion

Purpose of this project was to identify Toronto areas close to center with low number of restaurants (particularly Indian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Indian restaurant. By calculating restaurant density distribution from Foursquare data, we have first identified general boroughs that justify further analysis (Kensington Market and Cabbage Town), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighbourhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighbourhood etc.