

Capstone Project - The Battle of the Neighborhoods

- Prediction of an optimal location for a restaurant

1. Introduction

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 a French restaurant in Toronto, Canada.

1.1 Background

There are too many resto bars, coffeeshops etc. in the region. Therefore its is advantageous to use Data Science techniques to select an apt location in the region so that the business gets maximum returns.

1.2 Problem

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 the areas with no French restaurants in the vicinity.

We will use our data science powers to generate a few most promising neighborhoods based on this criteria.

Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

1.3 Interest

Specifically, this report will be targeted to stakeholders interested in opening a French restaurant in Toronto, Canada.

2. Data acquisition and cleaning

2.1 Data Sources

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

- Number of existing restaurants in the neighborhood (any type of restaurant)
- Number of and distance to French restaurants in the neighborhood.

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

- Wikipedia page:

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

- Geographical coordinates of each neighborhood
- Number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API.

For example, we can use the latitude and longitude of the neighborhood restaurants as the centroids and the area of interest say around 1km radius within the city.

2.2 Data cleaning

Data downloaded or scraped from multiple sources were combined into one table.

There were no outliers or missing values.

The data was filtered and used it for further analysis.

2.3 Feature selection

After data cleaning, there were 103 samples and 5 features in the data. Upon examining the meaning of each feature few common features were dropped.

:

	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
...
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944
99	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
100	M7Y	East Toronto	Business reply mail Processing Centre, South C...	43.662744	-79.321558
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999

103 rows × 5 columns

3. Methodology

In this project we will direct our efforts on detecting areas in central Toronto that have low restaurant density, particularly those with low number of French restaurants.

We will limit our analysis to areas within 1km in central Toronto.

In first step we have collected the required data: location and Neighbourhoods of Toronto.

Second step in our analysis we will explore the category and restaurant density ' across different areas of Central Toronto.

In third and final step we will focus on most promising areas and within those we will 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 1km.

We will create clusters (using ****k-means clustering****) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

4.Results section

Our analysis shows that although there are many restaurants in central Toronto (close to 170 in a radius of 1 km).

There are pockets of low restaurant density fairly such as Lawrence Park.

Highest concentration of type of restaurants was detected as Italian followed by Mexican, Japanese and Asian.

At present The Annex, North Midtown, Yorkville, Davisville and Summerhill West, Rathnelly, South Hill, Forest Hill SE neighbourhoods only have French restaurant.

Lawrence park does not have any restaurants near by so it would be good to open a French restaurant here.

5. Discussion

After directing our attention to central Toronto and nearby venues, we clustered to create zones of interest which contain greatest number of location candidates.

Purpose of this analysis was to only provide info on areas close to central Toronto but not crowded with existing restaurants.

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 taken into account and all other relevant conditions met.

6. Conclusion

purpose of this project was to identify central Toronto areas with low number of restaurants (particularly French restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new French restaurant. By calculating restaurant density distribution from Foursquare data we have first identified neighborhoods that justify further analysis 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 neighborhoods 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 neighborhood etc.