Report on the possible locations to set up new restaurant and the most popular category of restaurants in the city of Toronto Canada

Daniel Suhendro

## Abstract

There are many places with restaurants in Toronto, the capital city of Canada. Where the ideal spot to open a new restaurant and what is the right restaurant to set up? This is the question that is addressed in this report. This report is written for the client who wants to set up a new restaurant in Toronto. By reading this report, the client can get a better understanding of the number of restaurants in various neighbourhoods in Toronto as well as the restaurant categories to help in the client’s decision making.

There are five datasets that were used to answer the main question. They are 2016 Toronto census data, the Toronto Wards and Community Council Areas geographical data, the postal codes in Toronto, the postal code area coordinates and the restaurant venues dataset from Foursquare. The census data and geographical data were used to determine the community council area that has the highest population density. The area with the highest population density then was explored further for restaurants in 500 m radius around the coordinates of postal code areas.

The restaurant data were processed further using one hot encoding method, statistical count and sums. The final dataset consists of the name of the borough, name of the neighbourhood, the three most popular categories of restaurants and the number of restaurants in neighbourhoods having the same postal code. Finally the postal code areas were clustered together according to the range of the number of restaurants and this information is displayed on the map of Toronto. It is found that the neighbourhoods where restaurants are popular are Ryerson and Garden District. The most popular restaurant category was ‘restaurant’ assumed to serve Canadian cuisines. The client is advised to set up his restaurant here, serving Canadian cuisine. This report also has several alternative suggestions for consideration.

## Introduction

The aim of this report is to recommend the best location to open a new restaurant in the city of Toronto in Canada based on data from a popular venue search app: Foursquare. The background of this report is a request from a client who wants to move to Toronto, Ontario, Canada. He has run a successful restaurant business before and he wants to open a new restaurant in the city. His problem is he does not know where the right location or neighbourhood to set up his business is. He has consulted me as a data scientist to find out, using data, where are the good locations in Toronto to open up a new restaurant, and what type of restaurant is popular in Toronto. His problem then can be restated as a business question: ‘Can we find out what are the good locations for restaurants in Toronto and what are the popular types of restaurants there using data?’ Solving this problem will help the client not only choose the right location for his restaurant but also choose what kind of restaurant to be set up in that area. This report is written for the client to help him in his decision making.

## Data

There are five datasets that are going to be used in this project. The first data set is the 2016 Toronto Census data.[1] The second dataset is the geographical data of Toronto City Wards and Community Council Areas.[2] The third data set is the Toronto postal codes beginning with M from Wikipedia.[3] The fourth dataset is the latitudes and longitudes of the postal codes in Toronto.[4] The fifth data set is the resulting data from Foursquare search API (Application Programming Interface). The data that are relevant for answering the business question is the location data of popular restaurants as well as the category of the restaurants from the Foursquare search API. Before we collect and use that data, however, we must determine what neighbourhoods or areas in Toronto that is suitable for opening up restaurants. It can be assumed since Toronto is the capital city of Canada; there are many restaurants in the whole city. However, the distribution of restaurants will not be uniform; there will be areas with more restaurants and areas that have fewer restaurants, even none. The first hypothesis used in this report is that the number of restaurants is proportional to the population density. The area with the highest population density then is the best location for opening up a new restaurant.

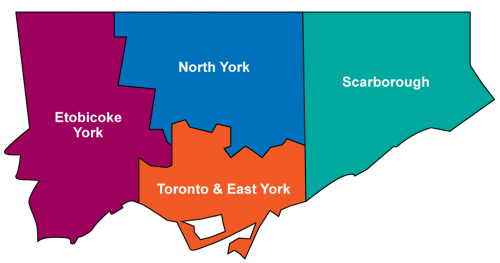


Figure the four community council areas (CCA) of Toronto, Canada. From[5].

The city of Toronto is divided into several levels of areas: community council areas (CCA), wards, and neighbourhoods. The CCAs are the first level of area division. It consists of several wards, which in turn consists of several neighbourhoods. There are four CCAs, North York, Etobicoke York, Toronto and East York, and Scarborough () with different population densities. To find the population density of each CCA, the 2016 Toronto Census data will be used. The census dataset consists of several categories such as total population per CCAs and the number of population in a certain age range. The geographical dataset provided the area size of each CCA. The population density is then calculated by dividing the total population of a CCA by the area size. For example, Etobicoke York has 642415 people [1] and an area 187.5 km2 [2] or 18750 hectares. The population density then will be 34 persons per hectare (rounded to nearest integer). Applying this method to all the CCAs, it is found that Toronto and East York has the highest population density: 80 persons per hectare. Since more population density means more potential customers this CCA is selected to be the area for the new restaurant.

Toronto and East York are divided into several postal code areas (PCA) and boroughs. Wikipedia provided the postal code data in the Toronto area, as well the corresponding boroughs and neighbourhoods in a table form [3]. The postal code in the city of Toronto begins with M. For example, in the column Postal Code, there is a code M1B. The Borough and Neighbourhood columns have Scarborough and Rouge respectively. A postal code can correspond to more than one neighbourhood in a borough. As an example, the code M1B also corresponds to the Malvern neighbourhood. This data will be used to divide the CCAs further to areas that have the same postal codes. Before use, the data needs to be pre-processed so that neighbourhoods that have the same postal code will be joined together in one row. This postal code data will be combined with the latitude and longitude of each postal code area.[4] The geospatial coordinates provided by the dataset consist of postal code column as well as latitude and longitude column. For example, the postal code M1B has the values 43.80669 and -79.1943534 in the corresponding latitude and longitude columns respectively. The combined data will provide a centre from which to search for popular restaurants. This data is finally going to be fed into the Foursquare search API to search for popular restaurants around the coordinates.

Foursquare is a service that enables travellers to search for popular sites and venues around their location [6]. This application also receives input and collects data about popular venues from users and so constantly updates its database. The database of venues in the Foursquare service can be searched by using Foursquare search API, a programming language that allows a programmer to search a specific kind of venues (in this case, restaurants) in an area defined by latitude, longitude and radius. In this project, the search for restaurant venues is limited to 100 venues in 500 m radius around the coordinates provided by the Wikipedia data. The result from the Foursquare search API is a JSON file consisting of restaurant name such as “Subway”, restaurant category such as “sandwich place”, restaurant address, restaurant website address if it has one as well as the restaurant latitude and longitude and some more information. The features that will be extracted from the JSON file will be the restaurant’s name, the restaurant’s category, and the restaurant’s latitude and longitude. These features will be combined with the dataset containing neighbourhood names, and the neighbourhoods’ latitude and longitude. The resulting dataset will consist of neighbourhood name, latitude and longitude, restaurant name, latitude and longitude, and the restaurant category. This resturant dataset will be explored for information on the number of restaurants in each PCA as well as the top three venue categories in order to answer the business question.

## Methodology

After the restaurant dataset has been prepared, the dataset was analysed for the number of restaurants in each PCAs. This was done by processing the dataset further so that the restaurants were grouped by the PCAs and the number of restaurants in a PCA was counted. The dataset was also sorted according to the number of restaurants to find out which PCA has the highest number of restaurants. In addition, the dataset was also analysed for the number of unique restaurant categories.

After the number of unique restaurant categories has been found, one hot encoding was performed on the dataset to classify the restaurants according to their category. One hot encoding is a method to turn data that have many categories into binary form [7]. For example the restaurant dataset has 59 categories of restaurants. The restaurants can be grouped into categories 1 to 59. By using one hot encoding, the dataset was transformed so that the restaurants of category 1 will have value 1 in the column of category 1 while other restaurants in the category of other than 1 will have value 0 in the category 1 column and so on for each of the categories. After each restaurant has been encoded, the restaurant categories were summed for each PCAs. This gives the number of restaurant of the same category in a given PCA. The dataset was then sorted further according to number of restaurants in each category for each PCA; this gives the popularity ranking of the restaurant venues in a PCA.

Finally the PCAs were clustered and labelled based on the number of restaurants in that area. Cluster 1 represented areas that have no restaurants, cluster 2 represented areas that have a few (between 1 - 9 ) restaurants, cluster 3 represented areas that have moderate amount of restaurants (between 10 to 30) and cluster 4 represented areas that have many restaurants (30 and more). Each cluster was represented with different colours on the map: red for cluster 1, purple for cluster 2, blue for cluster 3 and cyan for cluster 4.

## Results

Counting the number of restaurants shows that the Downtown Toronto borough has the highest number of restaurants (). Many of the PCAs in this borough were marked with blue or cyan representing restaurant numbers above 30 and in the range of 10 to 30 restaurants respectively ( and ). The neighbourhood that have the highest number of restaurants is Ryerson and Garden District which have 46 restaurants followed by Chinatown, Grange Park, and Kensington Market which have 45 restaurants.

There were 59 unique categories of restaurants in Toronto and East York area. The most popular restaurant category differs according to PCAs, but in general the order seems to be restaurants, followed by Chinese restaurants, Japanese restaurants and Italian restaurants. The map shows that most of the PCAs have 1-9 restaurants in an area within 500 m radius. The numbers of areas that have no restaurants and have more than 30 restaurants are the same. Finally 4 of the areas have restaurants that number in the range of 10 to 30. The areas that have more than 30 restaurants are located in the Downtown Toronto borough. also shows that for areas that have 14 restaurants or less the number of restaurants in each of the three most popular categories is the same which means at this point the ranking of popular restaurants cannot be taken as an indication of real popularity.

Table Complete data of restaurants in Toronto and East York. Each row represented a PCA, neighbourhoods which have the same postal code. The number inside the parentheses indicated the number of restaurant in that category.





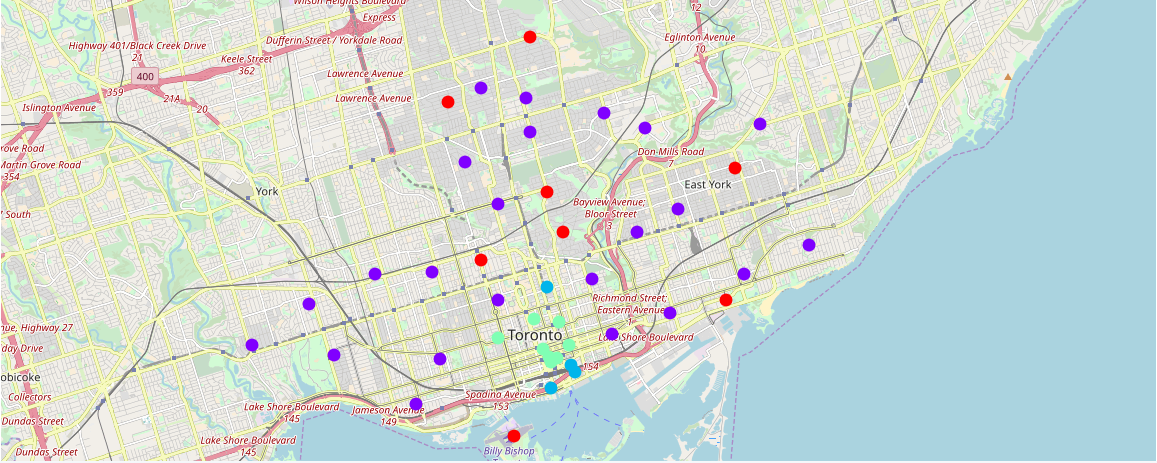


Figure map of Toronto and East York showing PCAs with restaurants. Red circles represent areas with 0 restaurants, purple represents areas with 1-9 restaurants, cyan represents 10-30 restaurants and blue represents 30 restaurants or more.

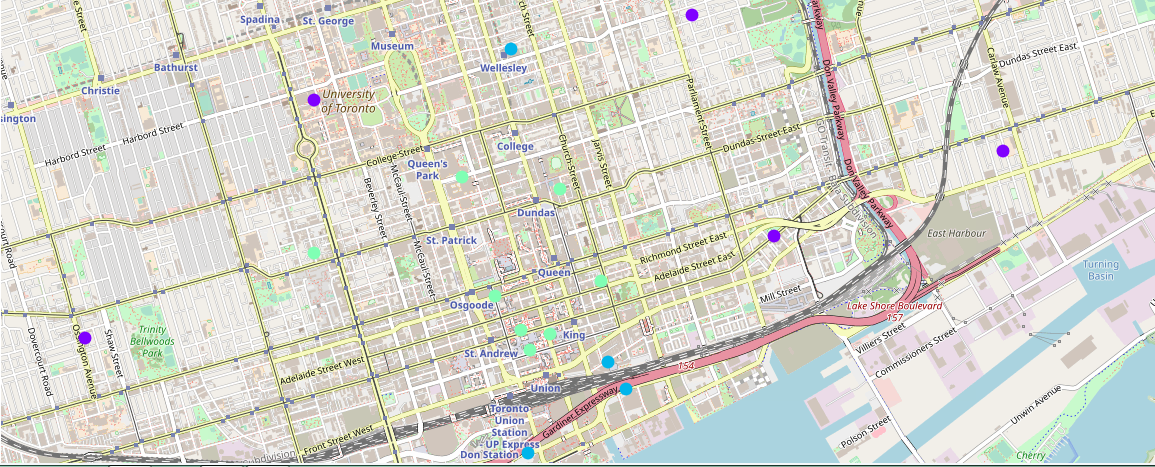


Figure Close up on the Downtown Toronto borough, showing the number of PCAs with 10 restaurants or more. It is apparent that there are 4 blue PCAs and 8 cyan PCAs.

## Discussion

The results of the data analysis shows that the places that have the highest number of restaurants are the Ryerson, Garden District, Chinatown, Grange Park, and Kensington Market neighbourhoods; all in the Downtown Toronto Borough. On the assumption that if an area has many restaurants, then that location is a good location to open up new restaurant, then these neighbourhoods are the best location for a new restaurant business. This answers the first part of the business question.

The most numerous type restaurants seem to be “restaurant”, which is assumed to serve Canadian cuisines since the country is Canada, followed by Chinese, Japanese and Italian restaurants. Assuming that the number of certain restaurant category in an area is proportional to its popularity, then the most popular type of restaurant venues is the “restaurant” category. Therefore it is recommended that the client set up restaurants in the five neighbourhoods mentioned and opening a restaurant that serves Canadian cuisines. This answered the second part of the business question.

Areas with high number of restaurants can also man high level of competition for a new restaurant business so an alternative suggestion to the client is that he may want to open up his restaurant business in cluster 3 or 2 areas (marked by blue and purple) such as Church and Wellesley in Downtown Toronto or Little Portugal in West Toronto which have lower competition than cluster 4 areas . Finally the red (cluster 1) areas such as Roselawn in Central Toronto and Rosedale in Downtown Toronto have no restaurant venues nearby so if the client decided to open his restaurant here he is guaranteed no competition but the risk of his business not succeeding is also high. Since these areas have no restaurants, further research need to be done to gather data on the feasibility of opening restaurants there, such as the eating out habit of the residents, the number of people that can work in a restaurants and most importantly, if the residents there want a restaurant in their neighbourhoods. These are out of scope for the current report.

## Conclusion

The data of restaurants have been analysed and the business question have been answered. The client has been recommended to open up his restaurant in Downtown Toronto choosing Canadian Cuisine as his restaurant’s specialty. There were some alternatives for the exact area for his new restaurant, depending on the level of competition. The area with highest number of restaurants is Ryerson and Garden District. This is the best area to open a new restaurant since these restaurant venues are very popular there. However these areas also have the highest competition of restaurant business. As another option the client can open his restaurant on areas that have 1 - 30 restaurants. Restaurant venues are slightly less popular but the competition is also reduced. Finally, areas with no restaurants and therefore no competition also exist in the Toronto and East York area, such as Rosedale and Roselawn so as the third alternative the client can try to open his restaurant here but more research needs to be done to determine the feasibility and future profits of opening up restaurants in such areas.

## References

1. *Toronto Census Data Ward Profiles*. 2016 [cited 2018 19 October ]; Available from: <https://www.toronto.ca/ext/open_data/catalog/data_set_files/Ward_Profiles_Census_2016.xlsx>.

2. *Toronto City Ward and Community Council Areas – Geographic Area*. 2016 [cited 2018 19 October]; Available from: <http://opendata.toronto.ca/it/com/Ward%20Profiles%20-%20WardAreas.xlsx>.

3. *List of Postal Codes of Canada: M*. 2018 [cited 2018 19 October]; Available from: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>.

4. Aklson, A. *Toronto Postal Code Areas' Latitudes and Longitudes*. 2018 [cited 2018 19 October]; Available from: <http://cocl.us/Geospatial_data>.

5. *Toronto Community Council Area Profiles*. 2018 [cited 2018 19 October]; Available from: <https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/community-council-area-profiles/>.

6. *Foursquare*. 2018 [cited 2018 19 October]; Available from: <https://en.wikipedia.org/wiki/Foursquare>.

7. *What is One Hot Encoding? Why And When do you have to use it?* 2018 [cited 2018 19 October ]; Available from: <https://hackernoon.com/what-is-one-hot-encoding-why-and-when-do-you-have-to-use-it-e3c6186d008f>.