# **Where to open a restaurant in Toronto?**

By Sebastian Wodjereck

**1. Description of the problem & Discussion of the Background**

Toronto is a diverse city with about 2.7 million inhabitants, while the Greater Toronto Area is home to a total of about 6.4 million inhabitants. Toronto encompasses a geographical area formerly administered by many separate municipalities. These municipalities have each developed a distinct history and identity over the years, and their names remain in common use among Torontonians. Throughout the city there exist hundreds of small neighborhoods and some larger neighborhoods covering a few square kilometers[[1]](#footnote-1).

As mentioned, Toronto’s neighborhoods are very different in terms of population density, but also in terms of “identity” and the type of people living there. Nevertheless, everybody wants to eat and go to a good restaurant in their neighborhood.

The goal of this study is to enable prospective restaurant owners to identify the neighborhood(s) where it is best to open a new restaurant. The recommendation will be based on the population density of each area, the restaurant density as well as the number of venues based on Foursquare data.

**2. Data Description**

For the analysis the following data will be used:

* Toronto’s neighborhoods based on postal codes as per Wikipedia (<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>). The data taken will be postal code, borough and neighborhood, e.g. M4A, North York, Victoria Village.
* Geospatial data for each neighborhood, i.e. their latitude and longitude value for each postal code ([http://cocl.us/Geospatial\_data](https://cocl.us/Geospatial_data))
* Population data for each of Toronto’s neighborhoods based on postal codes, based on the 2016 census of Canada (Source: Statistics Canada. 2017. Population and Dwelling Count Highlight Tables. 2016 Census). Here, the data used will be postal code and “Population 2016”, e.g. M4A, 14443. This means, a total of 14,443 people live in postal code area M4A.
* Data about restaurants in each neighborhood (number of restaurants and type of cuisine) taken from the Foursquare API (<https://developer.foursquare.com/developer/>). This step will involve a request for all venues in Toronto, by neighborhood based on their location data (latitude and longitude), but all venues other than restaurants will be dropped.

**3. Methodology**

For the work, I used Watson Studio, into which I imported the various data (i.e. neighborhood, geospatial coordinates, population and income) via csv upload, cleaned the data and merged them into a single dataframe of the following structure:

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung  
Table 1

I then imported all venue information on Toronto from Foursquare via the Foursquare API. First, I checked the information (i.e. if the venues were correctly imported and assigned to the corresponding neighborhood) by counting the number of venues in each neighborhood:

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung  
Table 2

Also, I checked how many venue categories there are in Toronto. The result is 270. Since we want to understand which neighborhood is best for opening a restaurant, I then further analyzed the results using one hot encoding, which resulted in the following table:

Ein Bild, das Tisch enthält.

Automatisch generierte BeschreibungTable 3

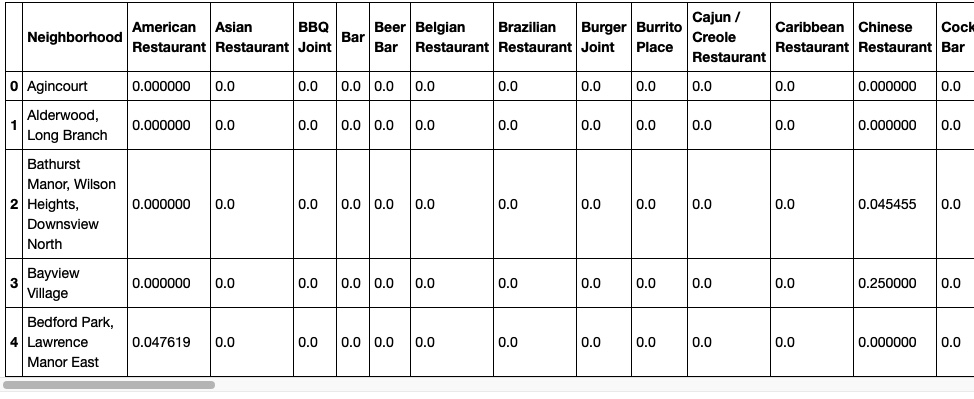
I then filtered out only those venues which represent a restaurant or some other kind of food place (i.e. Restaurant|Bar|Place|Joint):

Ein Bild, das Tisch enthält.

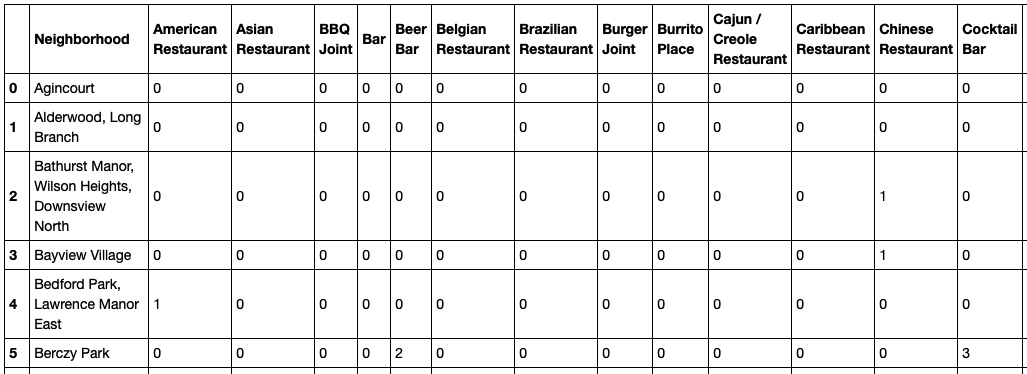
Automatisch generierte Beschreibung  
Table 4

In the following, I decided to group the data in two different ways:¶

1. by their means of occurrence for further analysis later on (in terms of a recommendation engine)

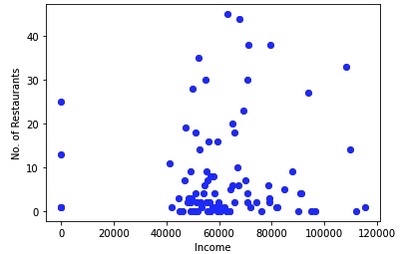
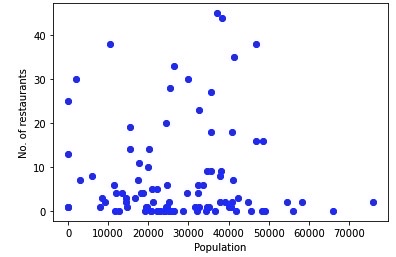
  
Table 5

1. by their total of restaurants to do some quantitative analysis

  
Table 6

In the Table 6, I also added a “Total” column, summing up the total number of restaurants in each neighborhood and also imported the population and income data from the first table mentioned in this section.

I then used that information to do some visualization of the data to determine if there is a correlation between the number of restaurants in a neighborhood and its average income as well as between its restaurant quantity and the neighborhood’s population. As can be seen in the following scatter plots, there is no linear correlation between either parameter and the number of restaurants in a neighborhood:

But since there is a bit of clustering around the middle in both, income and population, I decided to use each neighborhood’s weight in Toronto’s total population as well as average income as two of the three parameters for recommending where to open a restaurant.

For that purpose, I first deducted the means of the frequency of occurrence of a restaurant from 1, as the frequency of existing restaurants is inversely related to one wanting to open one there. I.e. if the frequency of occurrence in a given neighborhood is 10%, there is a higher chance of a successful opening of a new restaurant than in a neighborhood where the means of occurrence is already at 40%. In this example, the first neighborhood (with 10%) would get the weight 1 - 10% = 90%, while the second one (with 40%) would be scored lower with 1 – 40%= 60%.

I then merged Table 1 with Table 5 and added a column with the population and income weight (determined by dividing each respective neighborhood’s values for income and population by the total of these values across all neighborhoods). The weight of both, Population and Income do have a direct impact and thus are taken as is. All of this resulted in the following table, with the “Score” column being the product of columns “Total” (i.e. the inverse frequency of means of occurrence of restaurants in that neighborhood), Population Weight and Income Weight:

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung  
Table 7

**4. Results**

The final step was then to sort the Table 7 by “Score” and pull out the top 5 neighborhoods where it is best to open a restaurant:

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung  
Table 8

In addition, I visualized the five neighborhoods using a Folium map and the above-mentioned coordinates. The map will help those users who are not so familiar with Toronto to better locate the neighborhood they may favor.

Ein Bild, das Karte enthält.

Automatisch generierte Beschreibung

**5. Discussion**

As I mentioned before, Toronto is a big and diverse city. The population densities and income in each of the 97 districts can vary. As I showed in the scatter plots above, no one single attribute can be taken to determine what leads to the best location for opening a restaurant. Of course, many different methods can be used to determine that. In addition, information about a neighborhood’s ethnicity as well as whether it is rather a commercial or residential area would make the results more accurate.

I used various table operations to determine the best neighborhood to open a restaurant, taking into account the restaurant density in each area, its population and average income, giving each attribute a weight in coming up with a total score for each neighborhood. Obviously, this exercise can be expanded with additional attributes as mentioned above. Also, one could further go ahead and try and determine clusters (e.g. by using Kmeans) to further narrow down the choices.

**6. Conclusion**

Data analysis can help find a solution to a problem that at first glance seems impossible to solve. Like this, one can make a decision based on facts and figures and also based on visuals (e.g. if one decides to live in a certain area and doesn’t want to commute very far, it is easy to spot on the map).

1. https://en.wikipedia.org/wiki/Toronto#Neighbourhoods [↑](#footnote-ref-1)