Opening of new restaurants in Manhattan, NYC and Toronto: A comparative analysis

Contents

1 Introduction

Restaurants in metropolitan cities are not just places to dine, but also an opportunity for guests to experience new cultures via a variety of cuisines. Thus, it would be convenient to explore the districts where particular restaurants are located, when opening a new restaurant in a city. Incidentally, several factors contribute to the success of a new restaurant in a specific location [1]. We limit ourselves to the environment factor of the location, i.e. we try to find suitable locations with a high number of restaurants in their vicinity, since this may increase the number of potential customers [1]. We choose Manhattan, NYC and Toronto as the cities to perform the location data analysis project on, because both resemble each other in a few aspects. The business problem in this project maybe stated as follows: Can we find a suitable data analysis algorithm to predict whether it is more profitable to open a restaurant in either Manhattan, NYC or Toronto, and which neighbourhoods provide the best location? For this question we will also consider the kind of the restaurant in question, e.g. gourmet or street food, and the potential customers. The target audience in question are both gastronomical entrepreneurs and large and low scale investors, as the location data is useful for both small and big businesses. Since a restaurant's location can be crucial to its initial and long term success, location data analysis studies are crucial for restaurant owners and their investors.

2 Data acquisition

The following data will be used in the analysis of this project:

1. a json file conatining the geospatial data of New York City [2]
2. a list of postal codes for neighborhoods in Toronto taken from Wikipedia [3]
3. a csv file containing the longitude and latitude of the Toronto neighbourhoods [4]

The json file containing the geospatial data of New York City [2] is important in order to extract the Manhattan neighbourhoods alongside the latitude and longitude coordinates. In this way we can obtain the corresponding venues via a Foursquare

API "explore" query. In order to obtain the geospatial data of Toronto, we scrape a Wikipedia page [3] to obtain a table of postal codes and neighbourhood names of Toronto, using the Beautifulsoup package of Python. Then, we merge this table by the geospatial location data of each Toronto neighbourhood given in a csv file [4]. Finally, we can use a Foursquare API query to obtain the venues for each Toronto neighbourhood. For both the venues in Manhattan and in Toronto we will limit ourselves to "restaurant" category, since we are only interested in location data regarding restaurants.

3 Methodology

At first we would like to describe the data acquisition and data preparation for both the Manhattan and Toronto data sets.

3.1 The Manhattan data set

In order to obtain the Manhattan data set we first download a json file containing geospatial data of New York City [2] and store it into a pandas data frame. Then we create a new data set from the New York City data set containing only boroughs in Manhattan. We use the Python geolocator to obtain the geographical location of Manhattan and we use Foursquare API to get the 100 most popular venues within in 500 m radius. We group the data containing the neighbourhoods and their respective venues by the mean frequency of the venue category. Now we can employ an unsupervised machine learning algorithm, the so called k-means clustering to group the neighbourhoods and their venues into clusters with similar venue categories. Afterwards, we drop every venue category from each cluster which doesn't contain the word "Restaurant". We use the Python folium package to visualise the neighbourhood clusters on a map.

3.2 The Toronto data set

We scrape the Wikipedia page containing the postal codes, boroughs and neighbourhoods in Toronto [3] with the Python beatifulsoup package and store it into a pandas data frame. In this data set we drop non assigned boroughs, replace non assigned neighbourhoods with their respective borough names and group the data set by the postal code. Next we load a csv file containing the longitude and latitude of the Toronto neighbourhoods [4] and merge it with our panda data frame containing the Toronto neighbourhoods. We also restrict ourselves only to boroughs containing the word "Toronto". We use the Python geolocator to obtain the geographical location of Toronto and we use Foursquare API to get the 100 most popular venues within in 500 m radius. We group the data containing the neighbourhoods and their respective venues by the mean frequency of the venue category. Similarly to the Manhattan data set, we employ k-means clustering to group the neighbourhoods and their venues into clusters with similar venue categories. Afterwards, we drop every venue category from each cluster which doesn't contain the word "Restaurant". And once again, we use the Python folium package to visualise the neighbourhood clusters on a map.

4 Results

The results obtained via the k-means clustering are shown in the diagrams featuring clusters with a high, medium and low frequency of restaurant venues in both Manhattan and Toronto.

5 Discussion

The clustering for both Manhattan and Toronto produces rather interesting results. The neighbourhoods with a high density of restaurant are very dispersed in Toronto (large cluster 1 in red), but are much more dissected in Manhattan (cluster 1 in red, cluster 2 in purple and cluster 5 in orange). Furthermore, the Manhattan cluster 1 has

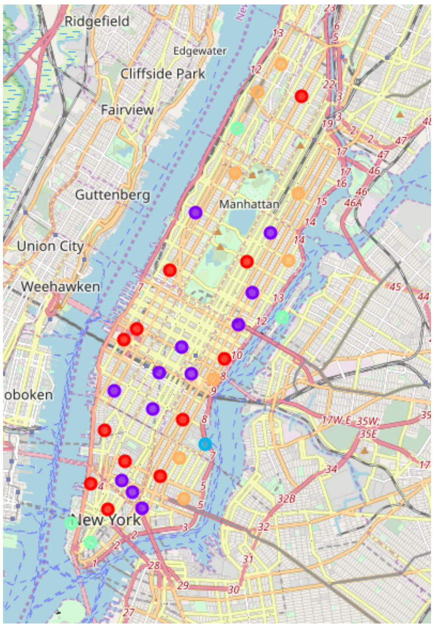

Figure 1: The neighbourhood clusters of Manhattan according to their most common venues.

	Neighborhood	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
0	Central Harlem	African Restaurant	Bar	American Restaurant	Seafood Restaurant	French Restaurant	Chinese Restaurant	Tapas Restaurant	Spa	Cosmetics Shop	Beer Bar
1	Upper East Side	Italian Restaurant	Exhibit	Coffee Shop	Bakery	Art Gallery	Juice Bar	Gym / Fitness Center	French Restaurant	American Restaurant	Hotel
2	Greenwich Village	Italian Restaurant	Clothing Store	Sushi Restaurant	Café	Indian Restaurant	Seafood Restaurant	French Restaurant	Dessert Shop	Gourmet Shop	Sandwich Place
3	Tribeca	American Restaurant	Park	Italian Restaurant	Spa	Café	Wine Shop	Wine Bar	Greek Restaurant	Men's Store	Coffee Shop
4	West Village	Italian Restaurant	New American Restaurant	Cosmetics Shop	Park	Cocktail Bar	Wine Bar	American Restaurant	Coffee Shop	Theater	Bakery
5	Noho	Italian Restaurant	Cocktail Bar	French Restaurant	Mexican Restaurant	Gift Shop	Bookstore	Rock Club	Coffee Shop	Pizza Place	Grocery Store
6	Civic Center	Italian Restaurant	Gym / Fitness Center	Coffee Shop	Sandwich Place	French Restaurant	Hotel	Yoga Studio	Cocktail Bar	Spa	Park
7	Turtle Bay	Italian Restaurant	Steakhouse	Sushi Restaurant	Coffee Shop	Wine Bar	Ramen Restaurant	French Restaurant	Park	Japanese Restaurant	Café
8	Hudson Yards	American Restaurant	Hotel	Gym / Fitness Center	Coffee Shop	Italian Restaurant	Café	Restaurant	Dog Run	Spanish Restaurant	Gym

Figure 2: Manhattan cluster 1 featuring only boroughs with restaurants as their 1st most common venue.

	Neighborhood	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
0	Chinatown	Chinese Restaurant	American Restaurant	Cocktail Bar	Salon / Barbershop	Spa	Optical Shop	Bakery	Vietnamese Restaurant	Hotpot Restaurant	Asian Restaurant
1	Upper West Side	Italian Restaurant	Coffee Shop	Wine Bar	Bar	Bakery	Mediterranean Restaurant	Café	Gym / Fitness Center	Ice Cream Shop	Indian Restaurant
2	Little Italy	Italian Restaurant	Café	Bubble Tea Shop	Bakery	Mediterranean Restaurant	Sandwich Place	Pizza Place	Cocktail Bar	Clothing Store	Salon / Barbershop
3	Midtown South	Korean Restaurant	Hotel	Hotel Bar	Japanese Restaurant	Coffee Shop	Dessert Shop	American Restaurant	Gym / Fitness Center	Cocktail Bar	Scenic Lookout

Figure 3: Manhattan cluster 2 featuring only boroughs with restaurants as their 1st most common venue.

	Neighborhood	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
0	Inwood	Mexican Restaurant	Lounge	Café	Pizza Place	Restaurant	Bakery	Park	Chinese Restaurant	American Restaurant	Frozen Yogurt Shop
1	East Harlem	Mexican Restaurant	Thai Restaurant	Bakery	Latin American Restaurant	Deli / Bodega	Cuban Restaurant	Pizza Place	Beer Bar	Taco Place	Gas Station
2	Yorkville	Italian Restaurant	Coffee Shop	Gym	Bar	Sushi Restaurant	Deli / Bodega	Bakery	Wine Shop	Diner	Japanese Restaurant

Figure 4: Manhattan cluster 5 featuring only boroughs with restaurants as their 1st most common venue.

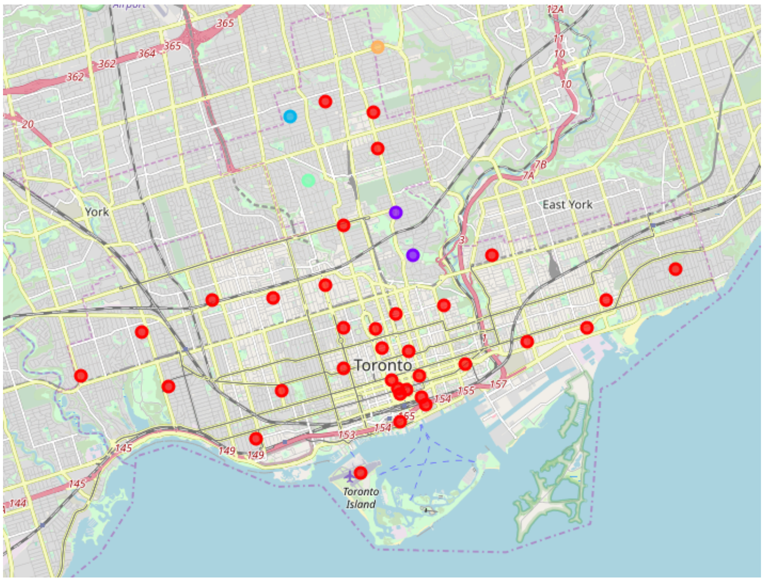

Figure 5: The neighbourhood clusters of Toronto according to their most common venues.

	Borough	Cluster Labels	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
0	East Toronto	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Furniture / Home Store	Ice Cream Shop	Dessert Shop	Bookstore	Brewery	Bubble Tea Shop	Restaurant
1	West Toronto	0	Thai Restaurant	Furniture / Home Store	Mexican Restaurant	Café	Bar	Grocery Store	Bookstore	Flea Market	Bakery	Cajun / Creole Restaurant

Figure 6: Toronto cluster 1 featuring only boroughs with restaurants as their 1st most common venue.

a very high number of top venues listed as restaurants and clusters 2 and 5 a medium number. We find that only the cluster 1 (red) in Toronto has a relatively low number of venues listed as restaurants. We may therefore conclude, that the density of top venue restaurants is significantly higher in Manhattan than in Toronto. However, it has to be mentioned, that the clustering method and data analysis performed here is very preliminary and can be refined by choosing an optimal number k of cluster for k-means clustering for instance.

6 Conclusion

We may conclude that the success of newly opened restaurants is predicted to be greater in Manhattan than in Toronto based on our data analysis performed here and entrepreneurial evidence [1]. Nevertheless, we should mention that having less restaurants in your neighbourhood can also be beneficial competition is lower. Moreover, we would like to stress the limitations of the data analysis performed here which used a rather simple k-means clustering algorithm, which would yield better results, if it was more refined.

References

[1] https://www.entrepreneur.com/slideshow/299849

[2] https://cocl.us/new_york_dataset

[3] https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

[4] https://cocl.us/Geospatial_data