



INDIAN RESTAURANTS IN TORONTO

Where to find one and where to open an Indian Restaurant

ABSTRACT

In this project we try to find the neighbourhoods with the maximum number of Indian restaurants and also attempt to find the best locations to open a new Indian Restaurant

Saurav Banerjee

IBM DATA SCIENCE CERTIFICATION

Contents

Introduction	2
Business Problem	2
Data Collection	2
Methodology	2
Determining the Neighbourhoods with maximum number of Indian Restaurants	3
Which Neighbourhood should you open an Indian Restaurant?	4
Analysing the clusters – Cluster 0	4
Analysing the clusters – Cluster 1	5
Analysing the clusters – Cluster 2	7
Analysing the clusters – Cluster 3	8
Analysing the clusters – Cluster 4	9
Analysing the clusters – Cluster 5	10
Final Recommendations	11
Conclusion	11

Introduction

Toronto is the capital of Ontario state in Canada and is the most populous city of Canada. Toronto is also a multicultural city with people from different cultures and ethnic backgrounds calling the city as their home. The diversity in culture also represents itself in the food and cuisine of Toronto, which results in some really great choice of food and venues. Culinary festivals and food tours of the city also play a role in popularizing the multicultural diversity of its cuisine.

Business Problem

There are a number of professionals who travel extensively across the major cities in the globe and finding the right food venue of your choice is important. This project will focus on addressing the following two requirements:

1. Determine the neighbourhoods which offers the maximum choice for Indian food in Toronto.
2. Which are the best neighbourhoods to open a new Indian Restaurant in Toronto.

While this project is focussed on the Indian cuisine, the same method can be applied to determine the best venues for other cuisines as well.

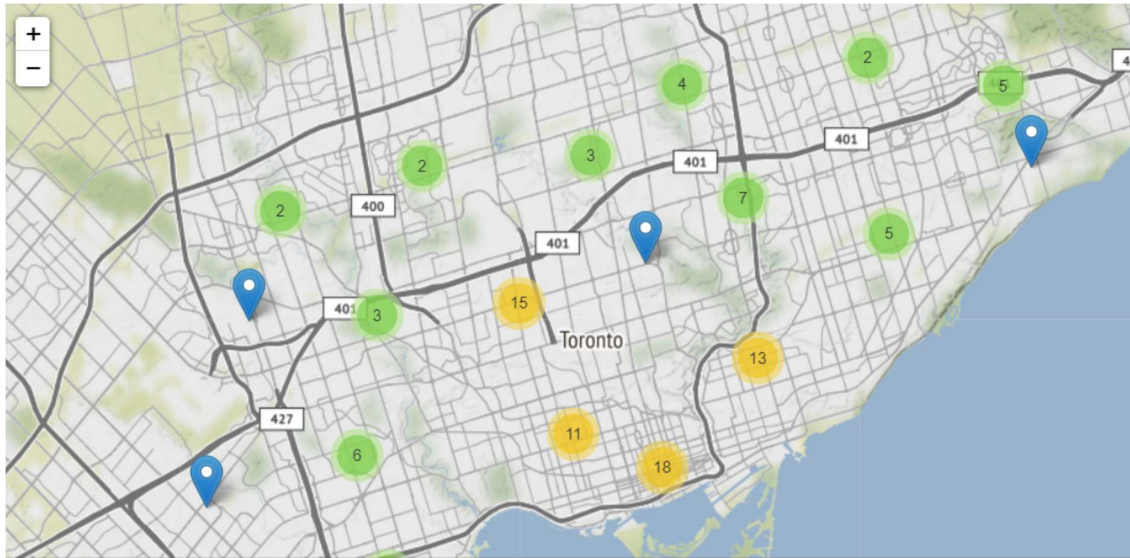
Data Collection

Data for the Toronto Community and the Neighbourhoods is collected from the following URL :
http://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

This data-frame is then combined with the latitude and longitude data for each Neighbourhood. Rows which did not have any value assigned for Community are dropped. Using the 'explore venue' call to the Four Square API, all venues near a neighbourhood are captured within a radius of 700. The venues are then merged with the Neighbourhood data frame to create the final data set which is used for analysis

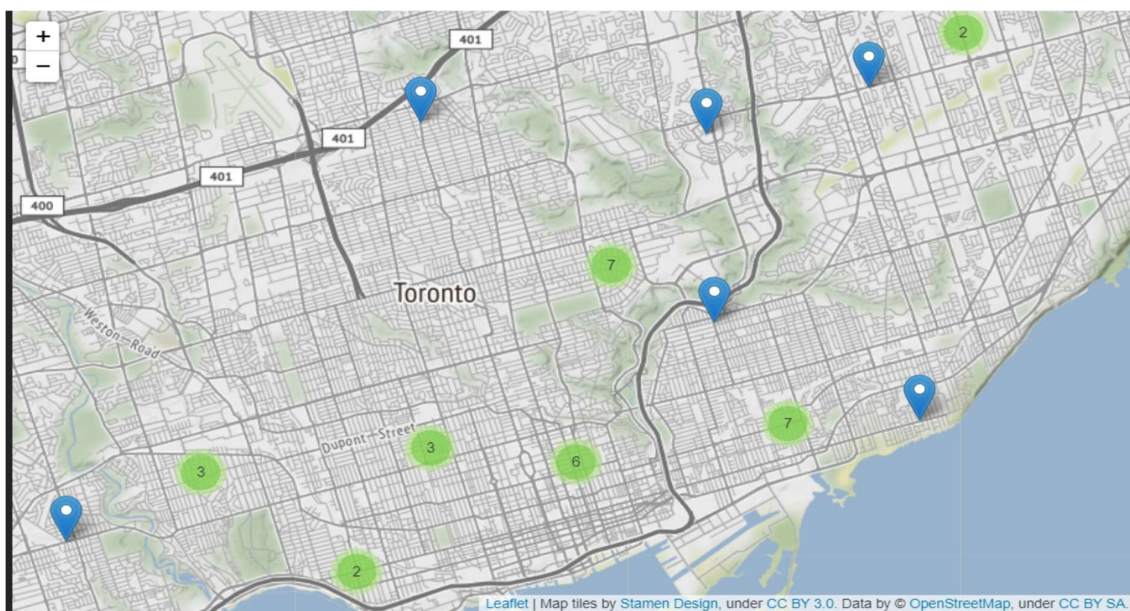
Methodology

Using the data frame which has the venues, neighbourhood, latitude and longitude data available we filter the dataset for which the 'Venue Category' is 'Indian Restaurant'. This provides a dataset which has 44 rows of data. Using this data set we then group neighbourhood to determine which neighbourhood has the highest number of 'Indian Restaurants'. The data for Indian Restaurants is then used to plot them on a map using the 'folium' library. Using cluster markers from the folium library we can create clusters of locations depicting the 'Indian Restaurants' to determine the locations in Toronto which offers the maximum number of Indian Restaurants in a Neighbourhood.



CLUSTER MAP OF ALL NEIGHBOURHOODS-FIGURE 1

Determining the Neighbourhoods with maximum number of Indian Restaurants



The Neighbourhoods with the highest number of Indian Restaurant is **'India Bazaar, The Beaches West'** which has 5 Indian Restaurants, followed by **'Thorncliffe Park'** which has 4 Indian Restaurants.

- There are 7 neighbourhoods which have 2 restaurants each, which are
 - Dorset Park, Wexford Heights, Scarborough Town Centre
 - Regent Park, Harbourfront
 - Christie

- Cedarbrae
- The Annex, North Midtown, Yorkville
- Canada Post Gateway Processing Centre
- Brockton, Parkdale Village, Exhibition Place
- Observing the clusters it is clear that ‘The Beaches West’ and ‘Thornccliffe Park’ are the best neighbourhoods to find a variety of Indian food given that they have the maximum options for Indian food.

Which Neighbourhood should you open an Indian Restaurant?

- In this part of the analysis we will cluster the neighbourhoods to determine the suitable neighbourhood which can be a potential location to open a new restaurant offering Indian food.
- Using K-Means clustering we create 5 neighbourhood clusters.
- To identify the neighbourhood cluster where it will be suitable to open a new restaurant, we use the following criteria
 - Should **have** some eatery in the top 10 venues for e.g coffee shop, pizza shop, restaurant.
 - May or may **not** have a Indian Restaurant in the top 10 venues. The reason for this is, if the cluster has a higher number of neighbourhoods and there are already Indian Restaurants in the top 10, then we can consider opening a Indian Restaurant in the remaining neighbourhoods in that cluster or decide against opening any restaurant. This will require further analysis which is not part of the scope of this project

Analysing the clusters – Cluster 0

The summary view of Cluster 0 along with the Top 2 Most Common Venues is listed below:

```
West Deane Park, Princess Gardens, Martin Grove, Islington, Cloverdale      1
Scarborough Village                                                         1
Alderwood, Long Branch                                                      1
Name: Neighborhood, dtype: int64
~~~~~
Ice Cream Shop                      1
Pizza Place                         1
Convenience Store                   1
Name: 1st Most Common Venue, dtype: int64
~~~~~
Convenience Store                    2
```

```
Pizza Place          1
Name: 2nd Most Common Venue, dtype: int64
~~~~~
```

- The first Cluster consists of 3 neighbourhoods.
- 1st Most Common Venue is: 'Ice Cream Shop'
- 2nd Most Common Venue is: 'Convenience Store'
- The number of neighbourhoods in this cluster is low and there are no restaurants in the top 2 most common venues.
- This cluster should not be considered for opening a new Indian restaurant as the number of neighbourhoods in this cluster is very low and therefore it will be a high risk if there are not enough customers for the new restaurant.

Analysing the clusters – Cluster 1

The summary view of Cluster 1 along with the Top 2 Most Common Venues is listed below:

```
Don Mills          2
North Toronto West, Lawrence Park
                  1
Stn A PO Boxes    1
Central Bay Street
                  1
Mimico NW, The Queensway West, South of Bloor, Kingsway Park South West, Royal York South West  1
..
Davisville North  1
Guildwood, Morningside, West Hill
                  1
Runnymede, Swansea
                  1
Rouge Hill, Port Union, Highland Creek
                  1
Queen's Park, Ontario Provincial Government
                  1
Name: Neighborhood, Length: 68, dtype: int64
~~~~~
Coffee Shop        23
Café                5
Indian Restaurant  4
Breakfast Spot     3
```

Fast Food Restaurant	2
Pizza Place	2
Hockey Arena	2
Italian Restaurant	2
Furniture / Home Store	2
Japanese Restaurant	2
Convenience Store	2
Home Service	1
Chinese Restaurant	1
Bakery	1
Rental Car Location	1
Brewery	1
Clothing Store	1
Park	1
Pool Hall	1
Hotel	1
Sporting Goods Shop	1
Miscellaneous Shop	1
Grocery Store	1
Discount Store	1
Bank	1
Pub	1
Dog Run	1
Greek Restaurant	1
Bar	1
Wings Joint	1
Intersection	1

Name: 1st Most Common Venue, dtype: int64

~~~~~

|                        |    |
|------------------------|----|
| Coffee Shop            | 12 |
| Café                   | 7  |
| Hotel                  | 6  |
| Park                   | 5  |
| Pizza Place            | 3  |
| Bar                    | 3  |
| Gym                    | 2  |
| Sandwich Place         | 2  |
| Japanese Restaurant    | 1  |
| Electronics Store      | 1  |
| Rental Car Location    | 1  |
| Yoga Studio            | 1  |
| Chinese Restaurant     | 1  |
| Mexican Restaurant     | 1  |
| Restaurant             | 1  |
| Spa                    | 1  |
| Italian Restaurant     | 1  |
| Ramen Restaurant       | 1  |
| Bakery                 | 1  |
| Athletics & Sports     | 1  |
| Furniture / Home Store | 1  |
| Clothing Store         | 1  |
| Fast Food Restaurant   | 1  |
| Health Food Store      | 1  |
| Pharmacy               | 1  |

|                            |   |
|----------------------------|---|
| Baseball Field             | 1 |
| Sculpture Garden           | 1 |
| Massage Studio             | 1 |
| Wine Shop                  | 1 |
| Grocery Store              | 1 |
| Tennis Court               | 1 |
| Hobby Shop                 | 1 |
| Construction & Landscaping | 1 |
| Sushi Restaurant           | 1 |
| Burger Joint               | 1 |
| Sporting Goods Shop        | 1 |
| Intersection               | 1 |

Name: 2nd Most Common Venue, dtype: int64

~~~~~

Cluster number 1 consists of 69 neighbourhoods.

- 1st Most Common Venue is : ‘Coffee Shop’
- 4 Neighbourhoods have ‘Indian Restaurant’ as the 1st Most Common Venue
- For this cluster we can consider opening an Indian Restaurant for the neighbourhoods which do not already have an Indian Restaurant as the 1st most common venue. However, this will require further analysis and is not part of the scope of this project

Analysing the clusters – Cluster 2

The summary view of Cluster 2 along with the Top 2 Most Common Venues is listed below:

North Park, Maple Leaf Park, Upwood Park	1
Woodbine Heights	1
Moore Park, Summerhill East	1
Caledonia-Fairbanks	1
Weston	1
Lawrence Park	1
Forest Hill North & West, Forest Hill Road Park	1
Birch Cliff, Cliffside West	1
York Mills West	1
Parkwoods	1

Name: Neighborhood, dtype: int64

~~~~~

|                      |   |
|----------------------|---|
| Park                 | 6 |
| Gym / Fitness Center | 1 |
| Gift Shop            | 1 |
| Bus Line             | 1 |
| Home Service         | 1 |

Name: 1st Most Common Venue, dtype: int64

~~~~~

Park	2
------	---


```

Diner      1
Convenience Store  1
Gym        1
Thai Restaurant  1
Food & Drink Shop  1
Grocery Store  1
Jewelry Store  1
Skating Rink  1
Name: 2nd Most Common Venue, dtype: int64
~~~~~

```

Cluster number 2 consists of 10 neighbourhoods.

- 1st and 2nd Most Common Venue is: 'Park'
- One Neighbourhood has 'Diner' and another 'Thai Restaurant' as the 2nd most common venue
- We can recommend opening a 'Indian Restaurant' for this cluster because:

—

Analysing the clusters – Cluster 3

The summary view of Cluster 3 along with the Top 2 Most Common Venues is listed below:

```

Willowdale, Newtonbrook  1
Woburn                   1
Rosedale                 1
Name: Neighborhood, dtype: int64
~~~~~
Park      2
Coffee Shop  1
Name: 1st Most Common Venue, dtype: int64
~~~~~
Trail     2
Park      1
Name: 2nd Most Common Venue, dtype: int64
~~~~~

```

Cluster number 3 consists of 3 neighbourhoods.

- 1st Most Common Venue is : 'Park'
- 2nd Most Common Venue is : 'Trail'
- We will **not** recommend opening a 'Indian Restaurant' for this cluster because of the low number of neighbourhoods and there are no restaurants in the top2 most common venues in this cluster.

Analysing the clusters – Cluster 4

The summary view of Cluster 4 along with the Top 2 Most Common Venues is listed below:

Downsview	4
Westmount	1
Steeles West, L'Amoreaux West	1
South Steeles, Silverstone, Humbergate, Jamestown, Mount Olive, Beaumont Heights, Thistletown, Albion Gardens	1
Parkview Hill, Woodbine Gardens	1
Clarks Corners, Tam O'Shanter, Sullivan	1
Willowdale, Willowdale West	1
Islington Avenue, Humber Valley Village	1
Milliken, Agincourt North, Steeles East, L'Amoreaux East	1
Kingsview Village, St. Phillips, Martin Grove Gardens, Richview Gardens	1
Bayview Village	1
Hillcrest Village	1
Name: Neighborhood, dtype: int64	
~~~~~	
Vietnamese Restaurant	4
Bank	2
Pharmacy	2
Chinese Restaurant	2
Pizza Place	2
Noodle House	1
Fast Food Restaurant	1
Intersection	1
Name: 1st Most Common Venue, dtype: int64	
~~~~~	
Grocery Store	5
Chinese Restaurant	3
Butcher	1
American Restaurant	1
Construction & Landscaping	1
Other Great Outdoors	1
Pizza Place	1
Coffee Shop	1
Shopping Mall	1
Name: 2nd Most Common Venue, dtype: int64	
~~~~~	

Cluster number 4 consists of 12 neighbourhoods.

- 1st Most Common Venue is: 'Vietnamese Restaurant'
  - For two neighbourhoods it is 'Chinese Restaurant'
  - For two other neighbourhoods it is 'Pizza Place'
- 2nd Most Common Venue is: 'Grocery Store'
  - For two of the neighbourhoods it is 'Chinese Restaurant'
- This cluster is ideally suited for opening a new 'Indian Restaurant' because
  - This cluster has the highest number of restaurants in the top 2 venues
  - There is no 'Indian Restaurant' currently in the top 2 venues, so opening an Indian Restaurant in this cluster will offer to the customers more variety of food.

## Analysing the clusters – Cluster 5

The summary view of Cluster 5 along with the Top 2 Most Common Venues is listed below:

```
Old Mill South, King's Mill Park, Sunnylea, Humber Bay, Mimico NE, The Queensway East, R
oyal York South East, Kingsway Park South East    1
Name: Neighborhood, dtype: int64
~~~~~
Construction & Landscaping 1
Name: 1st Most Common Venue, dtype: int64
~~~~~
Baseball Field    1
Name: 2nd Most Common Venue, dtype: int64
~~~~~
```

Cluster number 5 consists of 1 neighbourhood.

- 1st Most Common Venue is: 'Construction and Landscaping'
- 2nd Most Common Venue is: 'Baseball Field'
- We will not recommend opening a 'Indian Restaurant' for this cluster because there is only 1 neighbourhood and the top 2 most common venues are not restaurants.

## Final Recommendations

1. The maximum number of Indian Restaurants are in the neighbourhood of 'India Bazaar, The Beaches West'. For a travelling professional this neighbourhood offers the maximum number and choice of Indian restaurants. The next choice of neighbourhood for finding Indian food is 'Thornccliffe Park' which has 4 Indian restaurants.
2. To open a new restaurant in Toronto, Cluster 4 – which has 12 neighbourhoods has some of the best neighbourhoods to consider. Further analysis will be required to determine the exact location of the restaurant; however the above analysis provides a starting point to find the neighbourhoods which can be used to shortlist a suitable location for opening a new Indian Restaurant in Toronto.

## Conclusion

The analysis of the dataset was done using the folium package for visualizing the neighbourhoods and K-means algorithm to cluster the neighbourhoods. The neighbourhoods were grouped into 6 clusters and each cluster was then examined to determine the feasible locations for opening a new Indian restaurant. One of the parameters to be noted from the dataset is that Asian cuisine – Vietnamese, Chinese, Indian along with Pizza and Coffee Shops are some of the top visited food venues across all neighbourhoods. This is a favourable parameter for consideration for anyone who wants to open a Indian restaurant as this shows a affinity towards Asian cuisine. Further analysis can be done to sub segment a cluster and determine based on additional data collected on a more precise location which can be used to open a new restaurant.