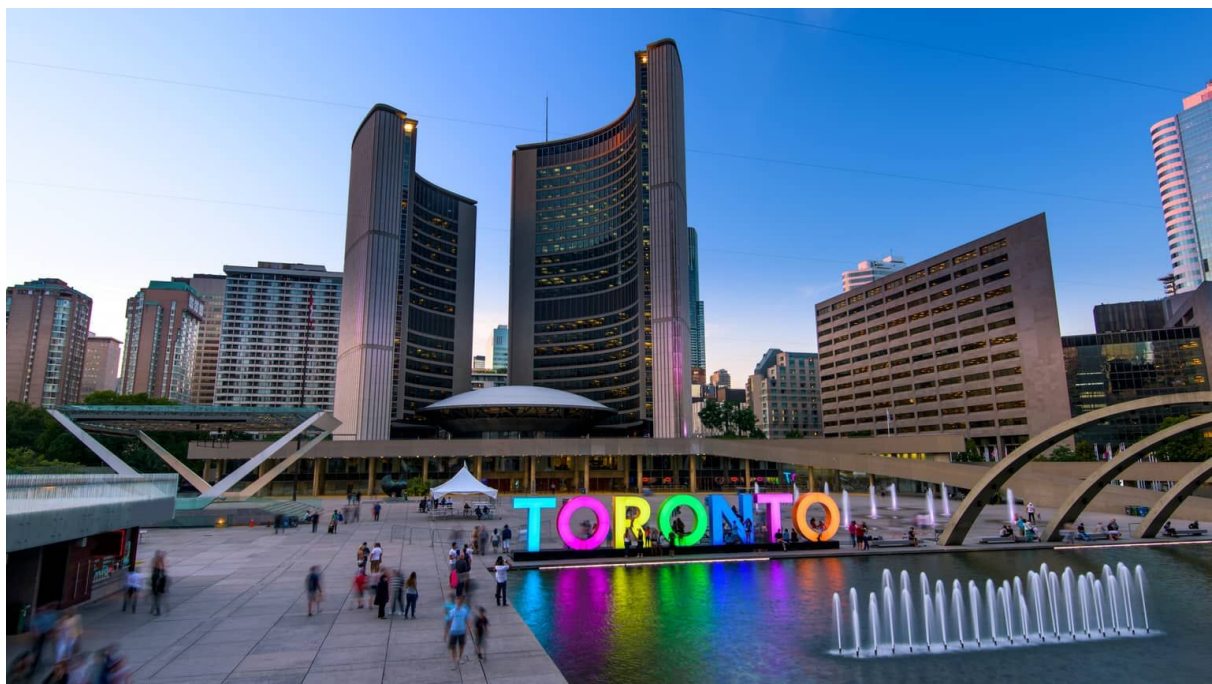


# IBM Data science project week 5th

## Introduction

This data science capstone project is for IBM Data science professional certificate. In this project I am creating a hypothetical situation for a concept that there may not be enough indian restaurants in Toronto. So it might be a great opportunity for a entrepreneur living in that area

As Indian food is popular among asian community, so this entrepreneur might be thinking of opening a restaurant in areas where the asian community resides. With the purpose in mind , finding the perfect location to open such a restaurant is one of the most important decision he has to take and i am designing this project to help him find best location for his restaurants



## Business problem

The Objective of this project is to find the best location for this entrepreneur to open a indian restaurant in Toronto, Canada. With the help data science methods and tools along with machine learning algorithm like clustering and foursquare API, this project aims to find the most suitable location to open a indian restaurant where most of the asian community resides

## Target audience

The entrepreneur who is looking for a suitable location in Toronto to open an authentic Indian restaurant

## Data section

To solve this problem, we will need the following data:

- List of neighborhoods in Toronto, Canada
- Longitude and latitude of these neighborhoods
- Venue data related to Indian restaurants. This data will help in selecting the most suitable neighborhood to open an Indian restaurant

## Data Extraction

- The scraping of Toronto neighborhoods via Wikipedia
- Getting latitude and longitude data of these neighborhoods using Geocoder package
- Using Foursquare API to get venue data related to these neighborhoods

## Methodology

- First I need a list of all neighborhoods in Toronto, Canada. This is possible by extracting the list of neighborhoods from Wikipedia:  
[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M.](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- Then, my next step was to clean the data and put the datasets in a dataframe
- Then, I need to get the coordinates of these neighborhoods to utilize Foursquare to pull the list of venues near these neighborhoods. To get the coordinates, I used the Geocoder package
- After gathering these coordinates, I visualize the map of New York using a Folium package to check the authenticity of these coordinates.
- Next, I use the Foursquare API to pull the list of top 100 venues within 500 metres radius. From Foursquare API I have been able to pull the names,

categories, longitude and latitude of these venues. With this data, I can check how many unique categories that i can get from these venues

- Then, I analyze each neighborhood by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category. This is to prepare for clustering to be done later
- Here I made a justification to specifically look for “indian restaurants”. Lastly I performed the clustering method by using k-means clustering algorithm. K-means clustering algorithm identifies k number of centroids and then allocates every data point to the nearest cluster while keeping the centroids as small as possible. I have clustered the neighborhoods in Toronto into 3 clusters based on their frequency of occurrence for “indian food”. Based on the results, i will be able to recommend a ideal location to the entrepreneur to open a restaurant

## Result

### CLUSTERS:



The results from k-means clustering show that we can categorize toronto neighborhoods in 3 clusters based on how many indian restaurants are in each neighborhood:

- Cluster 0: Neighborhoods with no indian restaurants
- Cluster 1: Neighborhoods with maximum number of indian restaurants
- Cluster 2: Neighborhoods with least number of indian restaurants

The results are visualized in the above map with Cluster 0 in red, Cluster 1 in blue and Cluster 2 in green

## Recommendation

- Most of the indian restaurants are in cluster 1 which is around St James Town, Church and Wellesley, Central Bay street, The Danforth West and Harbourfront East, lowest in cluster 2 areas which are in The Annex and Davisville
- There are good opportunities to open near St James Town, The Danforth West
- Looking at nearby venues it seems cluster 2 might be a good location as there are not much indian restaurants in that area and the competition is not fierce

## Conclusion

We can conclude from this project that cluster 1 has the maximum number of indian restaurants in that area and any new entrepreneur can select cluster 2 location to open a authentic indian restaurant