# **Capstone Project – The Battle of Neighborhoods**

## **Opening Indonesian Restaurants in Toronto**

## **By: Muhamad Fajrin Rasyid**

1. **Introduction**
   1. **Background**

Indonesian food is probably the tastiest invisible food in the world. It has been ranked highly by those who tried it, including recently by CNN, but many still do not know about it. It is not common to see Indonesian restaurant, much less than Thai or Japanese in many big cities around the world outside Indonesia. I suspect the same case is true for Toronto. Therefore, my hypothesis is that, if we develop the business of creating Indonesian restaurant there in the right way, it will become a successful business.

* 1. **Problem**

One of the biggest problems in creating restaurant is finding the right location. Where should we open our Indonesian restaurant? We might want to avoid area where there is already Indonesian restaurant (if any) or to some extent Chinese restaurant as it has slight similarity with Indonesian restaurant.

* 1. **Interest**

Hopefully this will bring benefit to anyone, especially entrepreneurs, who are interested in opening Indonesian restaurant business particularly in Toronto.

1. **Data acquisition and cleaning**
   1. **Data sources**

I am using below data to solve the problem above:

* List of neighborhoods in Toronto fetched from Wikipedia
* Neighborhoods latitude and longitude fetched from geocoder package
* Venue data especially related to Indonesian restaurant and Chinese restaurant fetched from Foursquare API
  1. **Data cleaning**

Before using the data, I need to do some preprocessing regarding the data below:

* Delete data where “Borough” column is not identified in the list of neighborhoods data which is fetched from Wikipedia
* Of the remaining data in the list of neighborhoods data, set “Neighborhood” column equals to “Borough” column if “Neighborhood” is not assigned
* Group neighborhoods having the same postal code
* Append the list of neighborhoods data above with its latitude and longitude which is fetched from geocoder package

The result of data preprocessing above can be seen in the screenshot below.

A screenshot of a cell phone

Description automatically generated

1. **Methodology**
   1. **Methodology summary**

In solving the problems above, I conduct steps below:

* Explore venues in each of the neighborhoods using Foursquare API
* From venues data, I try to specifically look for “Indonesian restaurant”
* I look as well for “Chinese restaurant” as it turns out that there is only one Indonesian restaurant in the whole Toronto
* I use K-Means to cluster neighborhoods into three clusters, those with many Indonesian or Chinese restaurants, those with some Indonesian or Chinese restaurants, and those with few or no Indonesian or Chinese Restaurants

In the next subsections I will explore each step taken above in more details.

* 1. **Explore venues**

I use Foursquare API to explore venues in each of the neighborhoods. To do this, I first define Foursquare credentials (CLIENT\_ID and CLIENT\_SECRET) as well as its version. I then define the function **“getNearbyVenues”** that will return all venues in each of the neighborhoods surrounding area in the radius of 500 meters. I limit the number of venues in each of the neighborhoods to 100 and store this data in **“toronto\_venues”** variable.

I then use the function shape and head to see examples of the venues above. It turns out that there are 2236 venues identified in toronto\_venues as seen below.

A screenshot of a cell phone

Description automatically generated

* 1. **Examine Indonesian restaurants**

I look for data having “Indonesian restaurant” as venue category as seen below.

A picture containing screenshot

Description automatically generated

Apparently, there is only 1 Indonesian Restaurant in Toronto! That makes it easy - we can just develop Indonesian Restaurant in any neighborhoods except one containing the current Indonesian restaurant. However, let us broaden the search to also include Chinese Restaurant as although not that similar, Indonesian food has slight similarity with Chinese food.

* 1. **Examine Chinese restaurants**

I look for data having “Chinese restaurant” as venue category as seen below.

A screenshot of a cell phone

Description automatically generated

It seems more promising – there are 23 Chinese restaurants in Toronto. We can then find which neighborhoods contain many Indonesian or Chinese restaurants, which contain some, and which contain few or none.

* 1. **Use K-Means to cluster neighborhoods into three clusters**

Finally, I use K-Means to cluster neighborhoods into three clusters:

* Cluster with neighborhoods which contain few or no Indonesian or Chinese restaurants
* Cluster with neighborhoods which contain some Indonesian or Chinese restaurants
* Cluster with neighborhoods which contain many Indonesian or Chinese restaurants

I first reshape the data to include each of the venues category as column as shown below.

A screenshot of a social media post

Description automatically generated

I then group rows by neighborhood and by taking the mean of the frequency of occurrence of each category and specifically choose "Indonesian Restaurant" and "Chinese Restaurant" columns. I group this into **“toronto\_grouped”** variable.

A screenshot of a cell phone

Description automatically generated

Here 0.0 means there is no Indonesian or Chinese restaurants in that neighborhood, and the higher number means the higher number of Indonesian or Chinese restaurants in that neighborhood.

I then ready to cluster the neighborhoods using K-Means on **toronto\_grouped** as shown below.

A screenshot of a cell phone

Description automatically generated

1. **Result**

As seen above, I have clustered neighborhoods into three clusters (cluster 0 with no or few Indonesian or Chinese restaurants, 1 with many, and 2 with some). I then merge **toronto\_grouped** with original **neighborhoods** data to get latitude and longitude of each neighborhoods as seen below.

A screenshot of a social media post

Description automatically generated

The merged data is enough for us to visualize the cluster in the map. I generate the map and below is the result. Note that red circle represents Cluster 0, purple represents Cluster 1, and green circle represents Cluster 2.

A close up of a map

Description automatically generated

We can now see which neighborhoods belong to which cluster below.

Cluster 0 - where there is no or few Indonesian or Chinese restaurants

A screenshot of a cell phone

Description automatically generated

Cluster 1 - where there are many Indonesian or Chinese restaurants

A screenshot of a cell phone

Description automatically generated

Cluster 2 - where there are some Indonesian or Chinese restaurants

A screenshot of a cell phone

Description automatically generated

1. **Conclusion**

Out of around 100 neighborhoods, only 5 neighborhoods have many Indonesian or Chinese restaurants, specifically those belong in Cluster 1, for example, Agincourt. More than 80% of the neighborhoods contain no or very few Indonesian or Chinese restaurants, specifically those belong in Cluster 0. Therefore, I recommend those who want to open Indonesian restaurant to start by looking areas in neighborhoods in Cluster 0.