



# CAPSTONE PROJECT - THE BATTLE OF THE NEIGHBORHOODS

Applied Data Science Capstone  
by IBM/Coursera





# CONTENT

INTRODUCTION

DATA AND  
EXTRACTION  
METHOD

METHODOLOGY

FINDING AND  
CONCLUSION





## CONTENT:

### **INTRODUCTION**

DATA AND  
EXTRACTION  
METHOD

METHODOLOGY

FINDING AND  
CONCLUSION

### ❖ INTRODUCTION

This project is to identify the best location for a Korean restaurant in Toronto, targeting for aspirational stakeholders that are interested to open a Korean cuisine restaurant in Toronto.

There may not be many Asian restaurant in Toronto, and specifically for Korean cuisine. However, we will try to detect potential locations that are not too crowded with restaurants and with no Korean restaurant in vicinity.

We will use data science methods and tools to assess of these criteria to answer the stated problem.





## CONTENT:

INTRODUCTION

**DATA AND  
EXTRACTION  
METHOD**

METHODOLOGY

FINDING AND  
CONCLUSION

### ❖ DATA AND EXTRACTION METHOD

To solve this problem, we will be using the following data and extraction method in *italics*:

- i. List of neighborhoods in Toronto, Canada  
*To be scrap from Toronto neighborhoods via Wikipedia*
- ii. Latitude and Longitude of these neighborhoods  
*To be obtained via Geocoder package*
- iii. Number of restaurants and their type and location in every neighborhood  
*To be obtained via Foursquare API*





## CONTENT:

INTRODUCTION

DATA AND  
EXTRACTION  
METHOD

**METHODOLOGY**

FINDING AND  
CONCLUSION

### ❖ METHODOLOGY

- i. Data pre-processing
  - The Toronto neighborhoods details were scrapped from Wikipedia and incomplete values ('not assigned') data are removed.
- ii. Using Geocoder package and obtained the geographical coordinates of the neighborhoods to be merged into the list extracted from Wikipedia.
- iii. Using Foursquare API to obtain names and category of the restaurant in each of the neighborhood. Next, we analyze each neighborhood to locate Korean restaurants.
- iv. Using k-means to cluster the neighborhood, we identified up to 3 clusters.



## CONTENT:

### INTRODUCTION

### DATA AND EXTRACTION METHOD

### METHODOLOGY

### FINDING AND CONCLUSION

#### ❖ FINDING AND CONCLUSION

	Neighborhood	Korean Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
4	Central Bay Street	0.014706	0	43.657952	-79.387383	Mo'Ramyun	43.656148	-79.392282	Korean Restaurant

	Neighborhood	Korean Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
19	Little Portugal, Trinity	0.022222	2	43.647927	-79.41975	OddSeoul	43.646192	-79.419601	Korean Restaurant

Based on the finding above, only 2 clusters show the presence for Korean restaurants which are Central Bay Street in cluster 0 and Little Portugal, Trinity in cluster 2.

No Korean restaurant found in cluster 1.

If the investor is interest to penetrate into a new market, then cluster 1 is a good choice.

Otherwise if the investor would want to open in an area with existing Korean cuisine customers then either cluster 0 or 2 can be considered.