

# The Battle of the Neighborhoods

A. Sethi

9<sup>th</sup> December 2019

## Introduction

In this project we are going to be analysing the city of Toronto and the neighborhoods within it. We are going to be analysing the Boroughs of Toronto and identifying the popular cuisines when it comes to the available food options within each of the Boroughs. The problem we are trying to solve here is to identify a potential location for a new restaurant and see which particular cuisines would provide a potential opportunity for success within those locations

## Dataset

The data set we will be using for this project comprises of the following.

We will be picking up neighborhood information 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).

We will be importing location, local business, popularity and other related information from the Foursquare API.

We will also be using a variety of python libraries.

## Methodology

We will be targeting a singular borough of Scarborough for this particular study. We will create a DF for the neighbourhood information for Scarborough, add lat. And long. Details to the same. Followed by the same, we would import venue information using the FourSquare API, group them by neighbourhood and sort the information using popularity metrics available via FS Api. We will further refine the information by classifying the businesses information in categories and then ordering them in order of popularity.

Our last step would be to cluster the businesses based on their geographic location into 3 specific clusters allowing us to develop further insight.

Step by step description of steps followed is provided below.

Import required libraries

Import Toronto Neighbourhood Data

Transform the data before we start using it

Create a Dataframe with the lat. and long. values of the neighbourhoods

Join neighborhood table with lat. and long. Information

Use geopy library to get the latitude and longitude values of Toronto, Canada

Create a map of Toronto with neighborhoods information markers

Segment and cluster only the targetted Borough in Toronto. In this case, that would be

Scarborough. So, let's slice and create a new dataframe of the Scarborough Neighborhood data

Use geopy library to get the latitude and longitude values of Scarborough, Toronto

Create a map of Scarborough with neighborhoods information markers

Define Foursquare credentials and version

Define function to explore neighbourhoods information

Refine results by getting information specifically for our targetted borough i.e. Scarborough

Group the venue results by categories

Analyze each Neighborhood

Group rows by neighborhood and by taking the mean of the frequency of occurrence of each category

Each neighborhood with the top 5 venues

Put the above information into a pandas dataframe

Create a new dataframe and display the top ten venues for each neighborhood

Cluster the Scarborough Neighborhood using k-means

Create a new dataframe that includes the cluster as well as the top ten venues for each neighborhood

Create a map using the above dataframe

Examine Scarborough Neighbourhood Clusters

Analyse the Scarborough Clusters 0, 1 and 2 to create insights and a conclusion

## Results

Based on the above information, we have identified the above conclusions

- Cluster 2 in Scarborough is clearly the largest with the largest audience size and hence, opportunity.
- There is not necessarily a singular cuisine or theme that can be identified to have an extremely strong or dominating presence.
- A fairly significant presence in top 10 popular venues for Chinese Restaurants, Indian Restaurants and Vietnamese Restaurants

## Conclusions

Based on the above results, we simply do not have enough conclusive data to identify an extremely dominant market space when it comes to the food industry. There are clearly strong players with a somewhat larger presence i.e. Chinese, Indian and Vietnamese restaurants. This can be concluded to provide one of two paths of decision making. Either the market can be considered saturated with these opportunities or it can be considered to be a safe option to invest in considering the large demand.

Overall though, our study has been somewhat inconclusive from a business opportunity research point of view.