**Capstone Project**

The Battle of Neighborhoods

**March 17, 2019**

1. **INTRODUCTION**

Back to few years ago Ipsos Reid along with the **Canadian Restaurant Food Association** did a survey about the frequency of eating out by Canadians and the reasons why reveals some interesting highlight such as 1) 63% of Canadians see eating out as a **luxury**, only 16% saw it as an everyday activity. 2) 34% of Canadians actually dream of opening up some form of food and/or drinking 3）23% of Canadians eat out once a week. 4) In the Atlantic and **Ontario, people on average eat out 2 times a week**.

The statista also revels there are 38% of Canadians eating outside 1~3 times per week in 2018 (<https://www.statista.com/statistics/561254/canada-average-fast-food-consumption-per-week/>) and many Canadians will continue to enjoy meals prepared outside of the home. Base on thess study, we can almost make a conclusion that running a restaurant business maybe a good idea.

But opening a restaurant requires juggling many moving parts and can feel like a difficult undertaking. One of the key task is finding a good Location and understand the competition of that area. Some nearby competition can help with marketing. But it’s wise to have enough of a distance that you can still guarantee a solid pool of customers who won’t be easily drawn to another similar place. So how should we decide the right location to open a high class Chinese restaurant in the Toronto city area?

1. **DATA DISCOVERY**
   1. **source of data & category**

To help the stakeholder decide a good location to open up the restaurant, the Toronto neighborhood location data (a general understanding of Toronto city neighborhoods from geo-location perspective) plus the neighborhood household income data (high class restaurant means target to higher income customers) will be used together to do the in-depth analysis and generate the Cluster to narrow down our target location selection. The Foursquare data (explore top 10 venues categories in each Toronto neighborhood) will also be used to explore the current Toronto restaurant locations to quickly understand the major competitors.

* Toronto neighborhood location data contains ‘Postal code’, ‘Borough’, ‘Neighborhood’ from the URL : <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M> and ‘Latitude and Longitude’ from URL : <http://cocl.us/Geospatial_data> but it did not work reliably, therefore I decided to use the csv file with postal codes provided
* Neighborhood household income data contains ‘Postal code’, ‘Borough’, ‘Neighborhood number’, ‘average household income’,…etc. from the URL : https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page\_Download-Telecharger.cfm?Lang=E&Tab=1&Geo1=CSD&Code1=3520005&Geo2=PR&Code2=01&Data=Count&SearchText=toronto&SearchType=Begins&SearchPR=01&B1=All&TABID=1
* Foursquare data contains ‘venue category’, ‘Latitude and Longitude’, …etc. using Foursquare open API to extract
  1. **data processing**

Toronto neighborhood location has been merged using the Postal code as join key to associate the neighborhood with Latitude and Longitude to form a new dataset (here we called ‘df1’)

This df1 will further merge with Neighborhood household income data using the Postal code as join key to form another new dataset (here we called ‘df2’)

Then I drop the unnecessary data columns including ‘Under $5,000',' $5,000 to $9,999', ' $10,000 to $14,999', ' $15,000 to $19,999', ' $20,000 to $24,999', ' $25,000 to $29,999', ' $30,000 to $34,999', ' $35,000 to $39,999', ' $40,000 to $44,999', ' $45,000 to $49,999', ' $50,000 to $59,999', ' $60,000 to $69,999', ' $70,000 to $79,999', ' $80,000 to $89,999', ' $90,000 to $99,999', ' $100,000 and over', ' $100,000 to $124,999', ' $125,000 to $149,999', ' $150,000 to $199,999', ' $200,000 and over', ' South Asian', ' Chinese', ' Black', ' Filipino', ' Latin American', ' Arab', ' Southeast Asian', ' West Asian', ' Korean', ' Japanese', 'White‘ then drop the NaN rows in order to fit into the cluster algorithm to produce the cluster labels in order to narrow down my target neighborhoods.

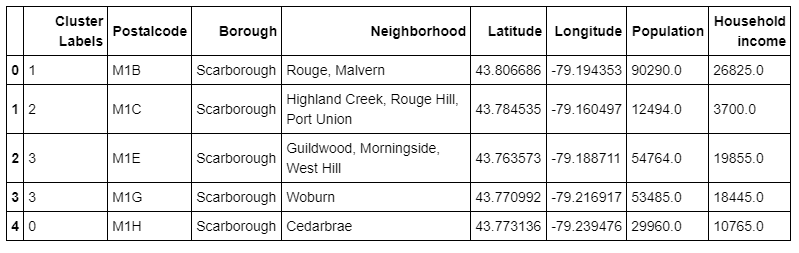
Client ID and Client Secret are prepared in advance through register the Foursquare developer account and using this Credentials (also Version) together with neighborhood latitude and longitude to extract venues related information then do the post processing such as ‘group by’, ‘sort’ and ‘one hot encoding’ to make it more readable.

1. **METHODOLOGY**
   1. **Clustering**

Leverage the unsupervised machine learning algorithm K-means to segment neighborhood into groups.

The machine learning algorithm divided the data into 5 groups and the ‘number of population’ and ‘household income’ are the major factors to come out the result set show as following Figure 1.

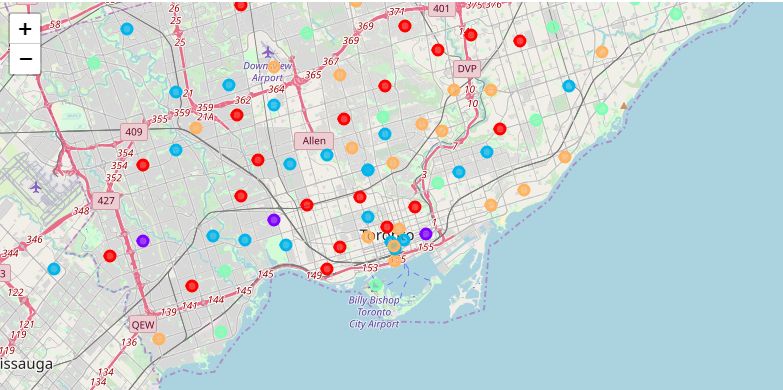
**Figure 1**



* 1. **Visualization**

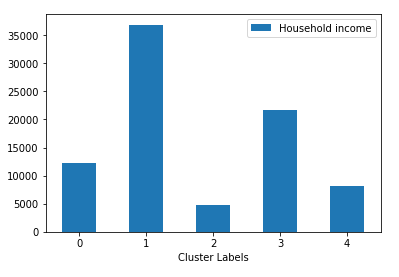
Leverage ‘folium’ (using “!conda install -c conda-forge folium=0.5.0 –yes” to install the package ) to draw the Toronto May (as Figure 2) allow me to explore the neighborhood insight from geo-location perspective. This is a useful technique especially for new comers who want to do the business with this big city without previous in-depth understanding.

**Figure 2**



Bar chart (as Figure 3) is also used to display the average household income group by cluster. That is extremely helpful for me to determine which cluster neighborhoods are my target to further analysis.

**Figure 3**



* 1. **Competitive Analysis**

In order to extract reasonable amount of data, the data limit equal to 200 was set pass to the Foursquare open API to extract the neighborhood nearby venue data.

The result set can be further massage to make it more readable format (Fig.4 and Fig.5) and also facilitate our intention to further determine which neighborhoods are the best of choice for opening a new Chinese restaurant.

**Figure 4**



**Figure 5**



1. **RESULT**

Cluster #1 was initially determined as target short list (Figure 6) of neighborhoods because 1) highest household income 2) highest population compare with the rest of 4 cluster neighborhoods

Agincourt North, L'Amoreaux East, Milliken, Newtonbrook Willowdale and Harbourfront Regent Park are de-seleced from our target because there are many competitors (Asian foods)

High Park The Junction South, Cloverdale Islington Martin Grove Princess Gardens West Deane Park and Rouge Malvern are ‘green light’ for stakeholder to do the ground survey to determine the final.

**Figure 6**



1. **DISCUSSION AND CONCLUSION**

Opening a restaurant is never an easy decision to make and needs lots of expert planning not only consider the right location but also do the following tasks right:

* **Choose a Restaurant Concept and Brand**
* **Form Your Menu Items**
* **a Restaurant Business Plan**
* **Obtain Funding**
* **Permits and Licenses**
* **Find an Equipment and Food Supplier**
* **Design a Restaurant Layout**
* **Hire the Right Staff**
* **Advertise Your Restaurant**
* **Host a Soft Opening**

However, as the restaurant industry continues to grow, and foodservice trends continue to diversify, there is always room for another extraordinary eatery. With detailed planning and successful execution of unique ideas, the restaurant business can flourish.

1. **APPENDIX**

* **Statista**

[**https://www.statista.com/**](https://www.statista.com/)

* **Canadian Restaurant Food Association** [**http://www.crfa.ca**](http://www.crfa.ca/)
* [**Wikipedia**](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

[**https://en.wikipedia.org/**](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

* **Statistics Canada**

[**https://www.statcan.gc.ca/eng/start**](https://www.statcan.gc.ca/eng/start)

* **Foursquare**

**https://foursquare.com/**