Capstone Project: Battle of the Neighbourhoods (Week 2)

Project Report By K Y Wong

A requirement of the IBM on Coursera Applied Data Science Course

Table of Contents

1. Introduction
2. Data
3. Methodology
4. Results
5. Discussion
6. Conclusion
7. Introduction - The Business Problem

A newly arrived immigrant plans to invest in a restaurant business in Toronto, Canada. The restaurant is to specialize in ethnic dining. The class of fine dining is not yet decided. Immigrant requests me as consultant to investigate and make recommendations. Another question to answer is the location of the restaurant.

Many factors will be considered, such as, the demographics of the location, economic well-being of the communities, the level of saturation with similar cuisines, the relative peacefulness (read crime intensity) and the draw of existing venues which ensure people traffic and patronage in the area.

Metro Toronto is growing and besides the old city centres and well-established hubs, other neighbouring communities growing beyond the metro area will be included for location selection as well.

Hence, the business questions to answer are where to locate the restaurant and the level of fine dining (type of service level and atmospheres) to aim for. Furthermore, how likely is the new restaurant in ethnic dining welcome to the neighbourhoods? Is it safe to do business in the neighbourhoods?

This business problem will interest other entrepreneurial immigrants, who just arrived to seek a better life for themselves and their families.

The study methodology and approach is applicable to the setting up of other types of amenities besides restaurant business as well.

The business problem definition and understanding will set the scene for analytic approach, data requirements and modelling.

The next section of the report will cover those aspects.

1. Data Section

In this section, the analytic approach is to gather data, prepare data and do data analysis, based on descriptive statistics and exploratory data visualization to understand the neighbourhoods in the Toronto, Ontario areas, the current popular venues and their amenities.

Clustering and analysing neighbourhood data gathered with Foursquare API endpoints and query for neighbourhoods in Toronto (Ontario province), Greater Toronto areas and not just metro Toronto will provide a larger geographical picture of opportunity.

The boroughs and neighbourhoods are extracted from the Canadian Postal Codes source: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:M> . Their corresponding latitudes and longitudes are checked off the data in the web link <https://cocl.us/Geospatial> data .

City of Toronto maintains a comprehensive database in the app “Wellbeing Toronto” (<http://map.toronto.ca/wellbeing/#eyJ0b3Itd2lkZ2V0LWNsYXNzYnJlYWsiOsSAcGVyY2VudE9wYWNpdHnElzcwfSwiY3VzxIJtYcSTYcSXxIBuZWlnaGJvdXJob29kc8S2fcSrxIHEg8SFxIfEicSLdGFixYXEmCLEo3RpdmVUxZBJZMSXxYnEhMWPYi1pbmRpY2HEgnLFhcWIYWdzTWFwxLYiesWCbcSXMTPErHjEly04ODM3NzYzLjXGhDcyN8SsxKc6NTQxMjkzMS4yNMaDMjg1xYjFpMWmxajFqsWSxIDFmMWraW9uxJcyxKxzxaRnbGXFtMSucsSTxJ9UaW1lxZzEqMSsxZbGucajIjfFtMafxafFqcSDxZxzQcWlV8S5xLt0xZJbxIDEh8WeOTQixKx3x41odMW5xKzEk8alc2VQb8SOcsSlxo5mYWzHoX1dxYfFiMa%2BZcehx6bHssWGxKzGs8a1dEnFpceFxapNxYPGsToixq1uxq%2FGscWH>) . This is used to access data on neighbourhood demographics: population, ethnic groups, family income and home prices in the neighbourhoods, business and economics. Wellbeing indicators measured by police/crime indicators also provide reference for neighbourhood safety. The database is based on census from 2014 or earlier(2008). It is used to construct categories of ‘Safe’, ‘Affluent’ and ‘ethnic presence/visibility’ in the neighbourhoods.

Comparisons between datasets in 2008 and 2014 (just released) on measures and indicators readily provides on further analysis as to whether the neighbourhood is vibrant and growing or decaying and in abandonment.

[www.Foursquare.com](http://www.Foursquare.com) and Foursquare API provide data about venues (amenities, restaurants, entertainment and social outlets, etc.) at neighbourhoods, their traffic/popularity and ratings by users and social commentators. Gives a profile of neighbourhoods’ relative penchant for types of facilities, amenities, restaurants, recreation, tourists attraction.

Modelling is based on unsupervised learning (i.e. clustering and segmentation of the neighbourhoods) to identify characteristics of clusters: ensure ‘fit’ choice for location or take considered risk to introduce novelty, innovation and uniqueness with a new business proposition in the location.

1. Methodology

Data imported from wellbeing Toronto need to be selected to infer the characteristics of ‘safe’, ‘affluent’ and ‘ethnic presence’ for the new restaurant to target menu, service and operate sustainably and profitably. They need to be cleaned and then statistics applied and explored using box-plotting, to provide quantitative markers for the three characteristics. Outliers found for family income and home prices (being indicators for ‘affluent’) are treated as a potential ‘super-affluent’ group, which could be targeted and offered level of service/dining as appropriate.

Data is cleaned and explored for family income distribution, unemployed rate, home prices distribution, crime rates ( robberies, break- in and entering occurrences) and presence of East Asians ethnic groups. Neighbourhoods are then filtered for ‘safe’, ‘affluent’ and ‘ethnic presence’ for selection of the restaurant location.

The chosen location(s) filtered will then be examined for the current trending traffic and popularity of venues, which will give some sense of similarity or fit of the business being developed to the chosen neighbourhood(s).

1. Results

The analysis of data and modelling of the neighbourhoods’ characteristics lead to the result that Banbury Don Mills (2014) now zoned as two neighbourhoods, Don Mills North and Don Mills South satisfy the criteria of ‘Safe’, ‘Affluent’ and with ‘Ethnic presence’ (East Asians – Chinese, Korean and Japanese.) They are therefore recommended as sites suitable for location of a new East Asian restaurant.

A Korean restaurant would be a novel type of cuisine to either location. A Chinese restaurant is better located in Don Mills North. Japanese restaurants are already popular in both locations. Both locations are popular for restaurants type businesses.

1. Discussion

Food appear to be of popular appeal in Don Mills North and Don Mills South. Popularity of fitness centres and restaurants in both locations suggest a healthy lifestyle type of food cuisine will be appealing. Perhaps explaining the popularity of Japanese restaurants at both Don Mills North and Don Mills South. Korean Hot Pots and Chinese steam-boat cuisines would have potential.

In view of the competitive state of Chinese and Japanese cuisines at both Don Mills North and Don Mills South, coupled with the general atmospheres of choice for dining, the potential for a fine dining style of Chinese cuisine is limited at these locations.

Perhaps a riskier option but worth further study is to locate fine dining Chinese or Japanese restaurant within the 500 m radius, closer to the park, garden and trail categories of Banbury-Don Mills neighbourhood for its exclusiveness.

1. Conclusions

The data science methodology and approach resulted in reaching the choice of location for the new ethnic restaurant with help of Apps like’ Foursquare API’ and open databases like ‘Wellbeing Toronto’ quite efficiently.

What this needs complementing by, is a more detailed input and evaluation of feed and first-hand information. Survey on the ground to validate the ‘non- analytic’ aspects like the ‘feel’ of the place or ‘buzz’.