Distribution Network of Craft Beer for a Start-Up

Eden Cheng

July 30, 2019

1. Introduction

1.1 Background

Our client is a startup craft beer brewer, they look for a distribution network of their craft beer in one of the area in Toronto. Since the supply of craft beer is limited, high target selling price, and the special flavor of the craft beer, they would like to find out the most suitable area to maximize their profit.

1.2 Challenge

There are some challenge that client would like to study, i) Craft beer with limited supply and "best tasting period", ii) High target selling price (i.e. 60% more expensive than branded beers e.g. Heineken, Budweiser) and iii) Special flavour like herbs, sours, salty lemon, etc. (Asian flavour)

1.3 Requirement from client

Client looks for an area with lots of bar/pubs/restaurants, to ensure that people in that area are willing to spend money on foods and drinks. Besides, area with Asian restaurants is preferred, as client thinks that it would be a selling point of craft beers with special flavor from Asia. Lastly, people are willing and affordable to spend money on the beers.

2. Data acquisition and cleaning

2.1 Data sources

Data could be found from below URL / API

- FourSquare developer API
- <List of postal codes of Canada>
 https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- <Latitude and Longitude data> http://cocl.us/Geospatial_data

2.2 Data cleaning

Data downloaded or scraped from multiple sources were combined into one table. There

were some of missing values from the source "List of postal codes of Canada". In this regard, I simply discarded the record without value for both the fields "Borough" and "Neighbourhood", or copy the value of "Borough" to "Neighbourhood". (see below)

Postcode +	Borough +	Neighbourhood	+		
M1A	Not assigned	Not assigned			
M2A	Not assigned	Not assigned			
мза	North York Parkwoods				
M4A	North York	rth York Victoria Village			
M5A	Downtown Toronto	to Harbourfront			
M5A	Downtown Toronto	owntown Toronto Regent Park			
M6A	North York Lawrence Heights				
M6A	North York	orth York Lawrence Manor			
M7A	Queen's Park>	Not assigned			
M8A	Not assigned——	Not assigned			
M9A	Etobicoke	Islington Avenue			
М1В	Scarborough	Rouge			

3. Data analysis

To form a set of data for analysis, I scraped the data from the URL <List of postal codes of Canada>, which contains Postal code, Borough and Neighbourhood. Besides, I downloaded the Latitude and Longitude data and merged the 2 datasets together, so that we could find out the Venues based on the location. (see below)

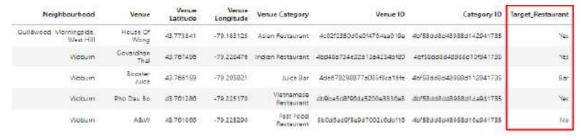
	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	MIE	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917

After that, based on the Latitude and Longitude, search nearby (i.e. radius = 1500m) Venue details of each "Neighbourhood" via FourSquare API, such as Venue name, Venue location, Category for further analysis. (see below)



Client looked for an area with lots of bar/pubs/restaurants with pricey menu, to ensure that people in that area are willing to spend money on foods and drinks, and preferable

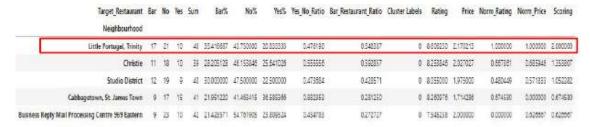
to have Asian restaurants as client thinks that it would be a selling point of craft beers with special flavor from Asia. Therefore, I filtered out the records with key words like Restaurant, Asian, Japanese, Bar, Lounge, etc to indicate the type of venue. (see below)



Further, I summarized the "Target Restaurant" and form a table which indicated the number of Bar, Asian Restaurant and Other Restaurant in each "Neighbourhood", calculated their percentage, and Bar-to-Restaurant & Asian-to-Other Restaurant Ratio. After that, I use K-Means Clustering method to categorize the Neighbourhood and confirmed to select Category 3 as our targeted areas for further study, as they have good Bar-to-Restaurant & Asian-to-Other Restaurant Ratio.



To prioritize the areas, we search extra venues information via FourSquare API, such as Rating and Price, to check if those Bar/Restaurant have good reputation and pricey menu. We got the average value of Rating and Price for each "Neighbourhood, and normalized them to come up a Scoring for our final judgement. (see below)



4. Conclusion

The study is based on geographical data scarped from URL and venue details via FourSquare API. By forming a set of relevant data, I selected relevant type of venues,

performed basic statistically analysis, and clustered them by using K-Means Clustering. With the help of venue details from FourSquare (i.e. rating and Price index), I could prioritize the selected areas and finally identified "Little Portugal, Trinity" is the best place for client to set up their distribution network of their Craft Beer. It is because there are many good rating bars and restaurants with pricey menu in this area, which implied a larger market. High ratio of Asian restaurant, which implied a higher chance that people accept the special flavour of Craft beer.

In case deeper study is required, we could extract extra data from FourSquare (e.g. menu, comment, etc) for a more sophisticated analysis.