Ideal location for clubs in Toronto

1. Introduction

We have a wealthy entrepreneur who wants to enter the business of nightlife: Clubs and bars. He wants to start with Toronto as base city and then proceed to setting up franchises in other cities in Canada too.

1.1. Problem

There are around 140 neighbourhoods in Toronto. The entrepreneur has to select the ideal area amongst these 140 neighbourhoods.

1.2. Target Audience

The target audience would be the potential investors willing to bet on our entrepreneur and his vision. From the analysis, we would be uncovering not only the potential neighbourhood but also what theme the club and bar should be to attract more customers.

2. Data acquisition and cleaning

2.1. Data Sources

Data has been sourced from Wikipedia as well as Toronto open database. Following links have been used:

Link to Toronto Data (age groups): https://open.toronto.ca/dataset/wellbeing-toronto-demographics/

Link to Toronto Data (income for neighborhoods): https://www.toronto.ca/wp-content/uploads/2017/12/9386-city-planning-tocore-neighbourhood-population-profiles-aoda-07-04-2016.pdf

Link to Canada Data (income for cities):
https://en.wikipedia.org/wiki/List_of_Median_household_income_of_cities_in_Canada

Link to Canada Population (population for cities):
https://en.wikipedia.org/wiki/List_of_the_100_largest_municipalities_in_Canada_by_population

2.2. Data cleaning

Data cleaning was required in all the data frames. The age group data was a excel file with 40 columns and 142 rows of data. Data required was for age group division for each neighbourhood. Additional columns for ethnic and language data was not needed and hence removed. Population for 2 years, 2008 and 2011 was available in the data set. Income data was required from median household income data. This dataset too had additional columns which were removed.

From Wiki, the data was scraped for Canada cities' population and income.

Foursquare API was used to get nearby bars and clubs of the preferred locations as well as trending venues in the area.

The population data for Toronto was available in digital format for only previous three census: 2008, 2011 and 2016.

3. Exploratory Data Analysis

3.1. Sorting and selection of neighbourhoods

The first step in data analysis was to set up the data for income and population for each neighbourhood. The data was cleaned and then merged, forming a data frame with 2008,2011 population and median household income, referred to as income.

The logic used is that a club owner would have to target the correct audience: The young adult age group; i.e.; Age group of 20-29 years. People of these age groups are generally more socializing and suited to atmosphere generally seen in clubs.

Moreover, a high income area means customers have high purchasing power. This directly implies their will to spend more in order to enjoy the lifestyles. This would mean better scope for profit margins on drinks and food being offered, a perfect recipe for high gains in this line of industry.

3.2. Narrowing down from neighbourhood options

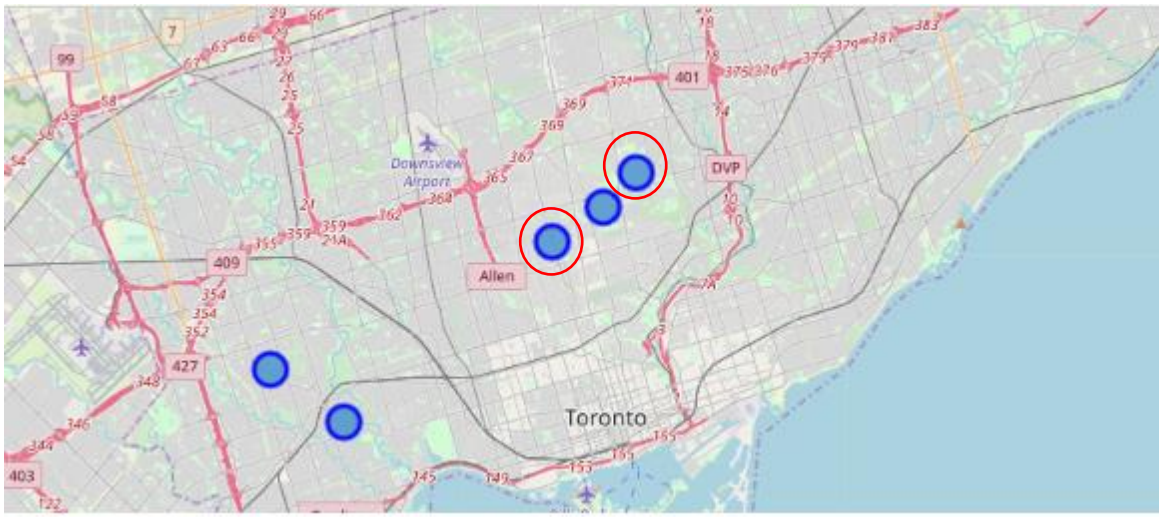
The data is sorted as per high income neighbourhoods. The top 5 high income neighbourhoods are selected from Toronto.

These neighbourhoods are then clustered on a map.

Neighbourhood must be so selected that it is close to all 5 top income neighbourhoods as well as has high population of young adults' age group. The neighbourhood must be appealing to people and theme must be set accordingly.

Once balance is obtained between above two criteria, the neighbourhood is finalized.

As per below map, two prime candidates can be identified:



The neighbourhood in north is Bridle Path-Sunnybrook-York Mills. This area is the most affluent in Toronto. The second choice is Lawrence Park South.

3.3. Finalizing the neighbourhood

The neighbourhood choice has been narrowed down to two neighbourhoods.

Now we need to check which area currently has more population of young adults' age group. Linear Regression has to be used to predict the population for 2019 using the population seen in 2008 and 2011 for the age groups.

Since we have only two previous data point available, linear regression in this case can be easily solved using point slope formula:

$$\text{Slope m} = \frac{y2-y1}{x2-x1}$$

Here y2 and y1 are y axis coordinates and x2 and x1 are x axis coordinates. We can take years on axis and population on y axis for determining the line equations.

Once slope of line is calculated, we can determine the predicted population using formula:

$$Y3 = y2 + m*(x3-x2)$$

This method is used to determine the predicted population of young adults in both narrowed down neighbourhoods, Lawrence Park South and Bridle Path-Sunnybrook-York Mills. Lawrence Park clearly has a much higher population in this age group.

3.4. Data analysis using Foursquare API

Now since we have finalized Lawrence Park South as ideal candidate, we need to check the geo data from Foursquare to confirm the neighbourhood′ s affluence, popular venues and nearby locations′ ratings so as to get an idea about popular opinion among its residents.

From trending places, The Granite Club is seen to be on the top for Lawrence Park South. The Granite club has facilities for most outdoor sports such as Lawn Tennis, Ice Hockey, Golf and others.

Foursquare also suggest presence of good rating bars and grills in the area, which shows people do enjoy an occasional drink in the neighbourhood.

A quick glance from Wiki for this area also suggests that most households have membership for The Granite Club. This suggests their inclination towards sports and thus a sports themed Club and bar would be a good way to kick-start business in this area.

4. Conclusions

We can conclude that Lawrence Park South would be a prime location for kick starting our wealthy entrepreneur′ s nightlife business. This neighbourhood offers affluent and young customers is abundance, a prime requirement in this industry. It is closest to cluster of high income neighbourhoods too. Moreover, the theme can be easily decided to be sports based since people residing in the area have interest in sports and this would attract them.

5. Future Directions

The Canada population and income data sets clearly point out two cities based on a logic similar to above: Montreal and Calgary. The potential investors can help in expanding the franchises to these cities.