

Selection of neighborhood for Club in Toronto

Data Science Capstone Project

Abhishek Rastogi

How important is neighborhood

- ▶ Clubs must target affluent customers who are have high purchasing power and thus are willing to spend more. This means a better profit margin and higher turnovers, a perfect recipe for kick-starting a business.
- ▶ The most common group of population visiting the clubs are young adults in age group of 20-29 years. The clubs must thus target the affluent people in this age group.
- ▶ The ideal neighborhood would be a combination of these target audience. The neighborhood should not only be affluent, with high young population, but also close to other affluent areas too.

Data acquisition and cleaning

- ▶ 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

Data acquisition and cleaning

- ▶ Data was cleaned since it contains several additional columns such as ethnicity, languages spoken and others. These additional data were dropped.
- ▶ Data was also scraped from Wikipedia.
- ▶ Foursquare API was used to find the nearby trending places and their ratings so as to determine the interest and affluence of the neighborhood.

Neighborhood Cluster



- ▶ Two highlighted neighborhoods, Lawrence Park South and Bridle Path-Sunnybrook-York Mills were the most suitable choices according to cluster.
- ▶ This cluster also shows that Lawrence Park South is area closest to the top 5 highest income areas. This makes it a prime candidate for us.

Linear Regression to understand the population metric of neighborhoods

- ▶ Linear Regression can be applied to calculate the predicted population of young adults for both the potential neighborhoods.
- ▶ Since only two previous data points are available, we can use the slope-line equation to find the predicted population for 2019.

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

$$\text{Slope } m = \frac{y_2 - y_1}{x_2 - x_1}$$

Here y_2 and y_1 are y axis coordinates and x_2 and x_1 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:

$$Y_3 = y_2 + m \cdot (x_3 - x_2)$$

Finalizing the neighborhood

- ▶ For Lawrence park south, the population of young adults is much higher than Bridle Path-Sunnybrook-York Mills.
- ▶ It is closest to all the 5 top income areas, including Bridle Path-Sunnybrook-York Mills which is the most affluent.
- ▶ Foursquare data shows presence of high rated bars and grills in the area. This shows population here enjoys an occasional drink which makes it a good choice for a club too.
- ▶ The Granite Club has several members from this area. It also offers good sports facilities.
- ▶ This makes the area the ideal candidate for a sports themed club and bar.

Conclusion and Future Directions

- ▶ 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 in abundance, a prime requirement in this industry. It is closest to a 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.
- ▶ 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.