Is a Childcare Facility in the Heart of the City a Viable Business?

Munish Kumar

Table of Contents

1. Introduction

Toronto is a busy, diverse city, and is one of the most populous cities in Canada. Toronto is also an international centre for business, finance, arts, and culture, and is recognized as being one of the most multicultural and cosmopolitan cities in the world.

According to Wikipedia, Toronto has a population of just under 3 million, with almost 50% being between the ages of 15-44. Also, 5% of the population is from 0-4 years of age. This is a fairly young population, with 10% of them having very young families (Wikipedia - Demographics of Toronto, n.d.).

It is also very likely that cost of living pressures in a city like Toronto would force either one or both parents to work. It is therefore no stretch of the imagination that a young population such as the one in Toronto, with its young families, will require access to childcare and early child education facilities, especially if there are busy working parents.

Often, such facilities are located in the suburbs, but what about the city centre, or somewhere close by? There is an opportunity space present to provide such a care facility, so would an enterprising individual/company consider opening a business related to early childhood somewhere close to the city? Would there be a demand for this business model? Would working parents appreciate such a facility?

2. Problem Statement

Determine if there exists the potential for a care facility aimed at young children and young families.

3. Targeted Audience

Businesses or individuals looking to branch into an early child care facility, nursery, pre-school or kindergarten.

4. Data Required & Sources

City Map of Toronto	Google maps
Suburb & Neighbourhood Information	Wikipedia or other sources
Geographical Coordinates	Wikipedia or other sources
Venues and Places in Toronto Area	FourSquares

5. Planned Analysis Steps

a) Steps undertaken in this analysis will include
b) Data Loading and Pre-processing. Data will be extracted using Beautiful Soup
c) Use the FourSquare API to download venue data and analyze for childcare facilities
d) Cluster the neighbourhoods using K-means and determine where, if any, childcare facilities are located and if there is a demand for said service

6. Methodology

a) The first step involves scrapping the data. In order to do this, I used beautiful soup, with Wikipedia as the data source (Wikipedia - Postal Codes, n.d.). I did the web scraping by utilizing pandas html table scraping method as it is easier and more convenient to pull tabular data directly from a web page into dataframe. This presented me with 3 columns of data, namely “Postal Code”, “Borough” and “Neighbourhood”.
b) This has to be next combined with postal code information; I used a data table from the following url ('http://cocl.us/Geospatial_data').

c) I remove some of the neighbourhoods that only 1 neighbourhood to reduce the size of the data set; in my analysis I ended up removing “Mississauga” only
d) I next use Folium to display the remaining neighbourhoods and the boroughs they are located in.
e) Next, I used the Foursquare API to generate a list of the top 100 venues, within a 800 meters radius of each neighbourhood latitude and longitude. In order to do this, I had to create a Foursquare developer account, which provided me with an account ID and API key to pull the data. From Foursquare, I am able to pull the names, categories, latitude and longitude of the venues. With this data, I can also check how many unique categories that I can get from these venues.
f) I now filter the venue category and only extract those venues of interest to me. In thos body of work, that would be neighbourhoods that have "Childcare Centre", "School", "Kindergarten" or "Pre-School" (“primary venues”). I create a new dataframe on which I perform one-hot encoding. I then cluster this data set.
g) Clustering is done using the K-means algorithm, which is an unsupervised machine learning algorithm. It works by identifying k number of centroids, and then allocates every data point to its nearest cluster based on Euclidean distance mapping, while keeping the centroids as small as possible. For simplicity, I have clustered the neighbourhoods into 3 unique clusters based on their frequency of occurrence for the venues of interest.

7. Complexities observed

After implementation of the steps expressed in (5) above, I discovered that the FourSquare API does not really show any of the “primary venues”. Rather it seems specifically skewed to places of interest, eateries, tourist attractions etc. This is perhaps what the website was designed for, and I simply misunderstood its categorization process. However, all is not lost as I figured that this presented a unique way to solve this challenge, in that I could then focus on “secondary venues” where childcare centres would thrive should the infrastructure be present.

Specifically, I began looking at places around which there existed such as a “Medical Center", "Playground", "Park", "Aquarium", "Soccer Field", "Swim School", "Summer Camp", "Museum", "Garden", "History Museum", "Office", "Baseball Field" and "Pool". While not all these venues are present (it’s a function of the radius), most were easily found by the API, and my logic is that, were these “secondary venues” close to a childcare facility, then an entrepreneur could entice customers by showcasing said venues. This might in fact strengthen the case for this business model.

The lack of data point was another thing I had to contend with, but perhaps this is not so surprising as such “secondary venues” cannot massively populate all neighbourhoods. After all, how many history museums can there be in a small radius?

8. Outcomes

The results of the k-means clustering are as follows:

9. Conclusion

There is indeed potential for this business of opening a childcare facility close to the city centre, but this must be carefully calibrated with other things. In my analysis, potentially, it seems like

somewhere within the “cluster 1” zone, there is scope to start one, given the existence of parks, gardens, museums, playgrounds and office infrastructure.

However, this analysis is only dependent on one data source i.e. the map data from Foursquare, as well as “offset data” from the “secondary venues”. An entrepreneur should actually scope the ground, and must also consider things like rent, availability of space, other competition in the form of nurseries not listed on the Foursquare site, and if there is a demand in that area in the form of young families.

Bibliography

Google Maps - Map of Toronto. (n.d.). Retrieved from https://www.google.com/maps/place/Toronto,+ON,+Canada/@43.7183953,-79.6582371,10z/data=!3m1!4b1!4m5!3m4!1s0x89d4cb90d7c63ba5:0x323555502ab4c477!8m2!3d43.653226!4d-79.3831843

Wikipedia - Demographics of Toronto. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Demographics_of_Toronto

Wikipedia - Postal Codes. (n.d.). Retrieved from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M