The Battle of the Neighbourhoods

Finding candidates among Toronto neighbourhoods to start a new private school

This report serves as a culmination of the IBM Data Science Professional Certificate course on Coursera. In this document, concepts and methodologies related to the entire course will be employed to solve a business problem.

Introduction

Toronto is the largest city in the most populous province of Canada. It is home to a large population of immigrants who come from all the four corners of the world, in addition to a large indigenous population. This leads to many differing demands for food choices, retail experiences and schooling needs.

The education system in Canada is one of the world's best but even the public education system cannot keep up with the myriad tastes, standards and expectations of parents and caregivers. This creates an opportunity to fill a demand to provide quality education catered to the students needs by way of private education.

Problem Statement

The large multicultural population in Toronto requires a more flexible and varied education system that goes above and beyond the one-size fits all public education system. This creates a business need for schools that offer more flexibilities in terms of subject choices and curriculum.

This proposal addresses this by proposing the setting up of a new private school. As many private schools already exist in the city, we shall use data science and machine learning to help us locate candidates among Toronto neighbourhoods for the new school.

Selection Criteria

Several factors will be considered when exploring possible neighbourhoods to place this new school.

1) Population Density

A new school will require a certain minimum number of students to make it a viable investment. Toronto neighbourhoods will be queried to obtain the population numbers and only areas above a certain level will be considered.

2) Number of existing private schools

Existing private schools will provide competition and affect the number of students and revenue to a new school. Thus neighbourhoods with many private schools will be excluded.

Assumptions

The following assumptions are taken when considering the data available and for reasons of expediency.

1) Religious schools will be considered public schools.
2) The number of public schools is assumed to have negligible impact (as they cater to different segments of the population).
3) The makeup of the population in the neighbourhood (race, religion, social standing etc.) will not be a factor.

4) The population breakdown across ages will be assumed to be fairly consistent across all neighbourhoods

Data Sources

Data from various sources will be used in our analysis.

1) Wikipedia page of Toronto Neighborhoods.

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

This gives us information on the boroughs, neighbourhoods and the postal codes. The postal codes, via their geolocation data, will allow us to search Foursquare for information on private schools in a particular neighbourhood.

2) Geolocation data on Toronto neighbourhoods.

http://cocl.us/Geospatial_data

This link contains a CSV file with the latitude and longitude information of Toronto neighbourhoods.

3) Foursquare location data to obtain information on private schools in Toronto neighbourhoods.

https://developer.foursquare.com/docs/build-with-foursquare/categories/

The Foursquare API search will use the venue categories for Private School **52e81612bcbc57f1066b7a46.** This will return a list of all the private schools within a given location.

4) Information on the population of each Toronto neighbourhood.

https://ckan0.cf.opendata.inter.prod-toronto.ca/download_resource/ef0239b1-832b-4d0b-a1f3-4153e53b189e?format=csv

The city of Toronto Open Data Portal has information on neighbourhood profiles based on the 2016 Population census by Statistics Canada.