Identifying the Gaps in University Coverage in Toronto

Tim Cayanga

March 25, 2020

1. Introduction

1.1. Background

Toronto has been consistently ranked as one of the top student cities in the world (11th in 2019, 13th in 2018, 11th in 2017)[1]. There are four universities in Toronto, namely Ryerson University, University of Toronto, York University, and Trent University[2]. Data shows that there are around 180,000 students studying at these universities[3]. Of the four universities mentioned, only Trent University is situated outside the city of Toronto.

In 2016, statistics show that 54% of Canadians aged 25 to 64 have a college or university diploma, which ranks first among the Organisation for Economic Co-Operation and Development (OECD) countries[4]. Census data shows that there are 245,605 people aged 18 to 25 in Toronto[5]. Multiplying the percentage of Canadians in college with the number of potential students in Toronto, there are roughly 132,000 people in college and 135,000 people who are not in college.

1.2. Problem

Ideally 100% of people would have the option of tertiary education, but they may be constrained by factors such as cost or distance. This project aims to determine the ideal locations

[1] https://www.topuniversities.com/city-rankings/2019
[2] https://studyabroad.shiksha.com/top-universities-in-toronto-articlepage-1533
[3] https://ocul.on.ca/populations
[4] https://www150.statcan.gc.ca/n1/daily-quotidien/171129/dq171129a-eng.htm
[5] https://www12.statcan.gc.ca/census-recensement/2011/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CSD&Code1=3520005&Geo2=PR&Code2=35&Data=Count&SearchText=Toronto&SearchType=Begins&SearchPR=01&B1=All&GeoLevel=PR&GeoCode=3520005

for new universities or satellite campuses in Toronto that serve the neighborhoods which may be too far from the existing universities while accounting for the total population of neighborhoods[6].

1.3. Interest

It would be of interest to Ontario's Ministry of Training, Colleges, and Universities to see in what areas of Toronto – the most populous city in Ontario – have limited access to universities, and therefore the province can create new public universities in the target areas. It would also be of interest to private universities who are interested in creating satellite campuses in Toronto to serve a greater population of students. Finally, it would be of interest to the youth of Toronto who may be looking to take up tertiary education but the current universities may be far away from them.

2. Data Acquisition and Cleaning

2.1. Data Sources

FourSquare, a technological company which uses a massive database of accurate location data, will be utilized to determine the locations and distances of universities in Toronto. Data for the total population of neighborhoods will be sourced from the census of Statistics Canada.

A post on medium.com by A Gordon was also used for information on the boundaries of Toronto neighborhoods and conversion to a Choropleth map[7].

2.2. Data Cleaning

The total population of each postal code was first tallied and placed in a dataframe.

[6] https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Table.cfm?Lang=Eng&T=1201&SR=1&S=22&O=A&RPP=9999&PR=0
[7] https://medium.com/dataexplorations/generating-geojson-file-for-toronto-fsas-9b478a059f04

Data collected from FourSquare will be compiled and will only take into account the three universities mentioned and any satellite campuses they have in Toronto. Other universities and colleges that are not recognized but are shown on FourSquare will be removed.

A problem was encountered when trying to determine the universities near Toronto, wherein most of the results shown were within 1,000 meters from the center of Toronto whereas York University was more than 16,000 meters from the center. As such, a separate search was performed to include York University.

2.3. Feature Selection

After data cleaning, there were 103 neighborhoods and 6 features, namely Postal Code, Borough, Neighborhood, Average Income, Latitude, and Longitude.

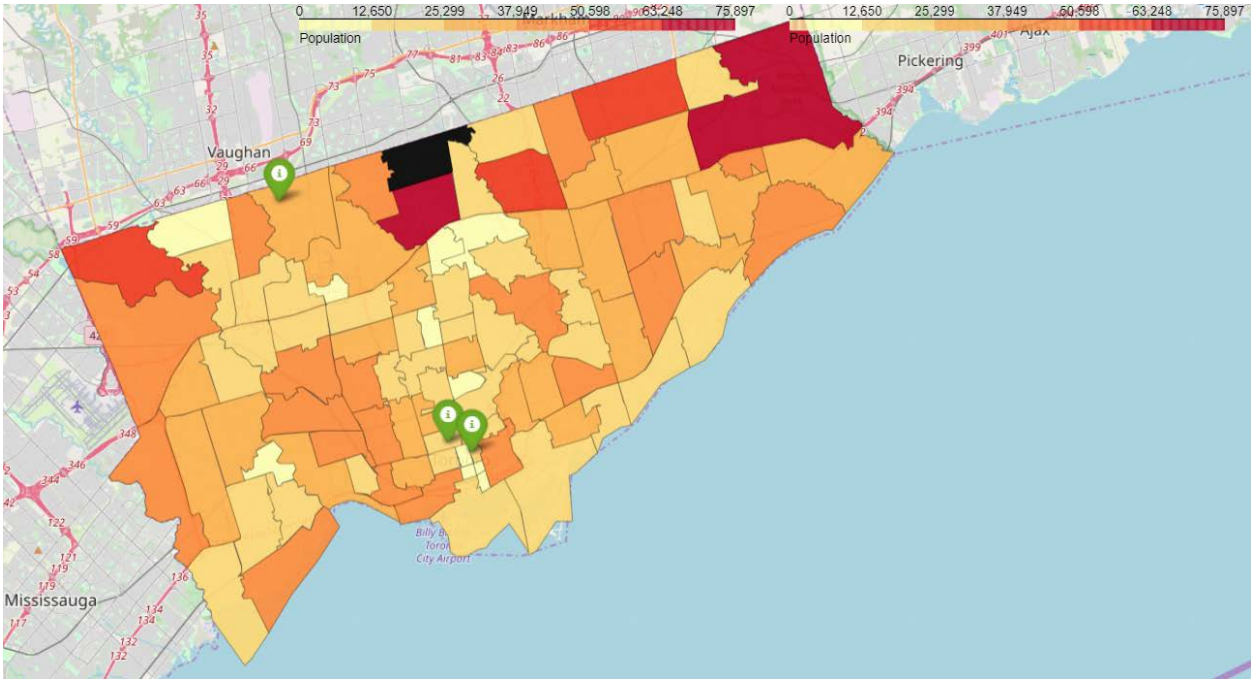
	Postal Code	Borough	Neighborhood	Total Population	Postal Code	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	66108	M1B	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	35626	M1C	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	46943	M1E	43.763573	-79.188711
3	M1G	Scarborough	Woburn	29690	M1G	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	24383	M1H	43.773136	-79.239476

The three universities contained 16 features, namely Name, Categories, Address CC (Country Code), City, Country, Cross Street, Distance from Toronto center, LabeledLatLngs, Latitude, Longitude, Neighborhood, Postal Code, Province, and ID.

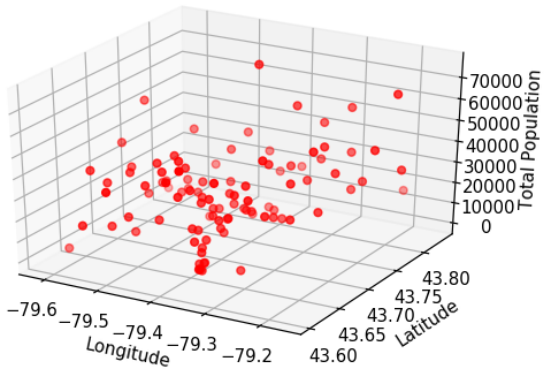
	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng	neighborhood	postalCode	state	id
0	University of Toronto	University	27 King's College Cir	CA	Toronto	Canada	at King's College Rd	1148	[27 King's College Cir (at King's College Rd),...	[{'label': 'display', 'lat': 43.6624934706167,...	43.662493	-79.39522	NaN	M5S 1A1	ON	4ad4c05ef964a52097f620e3
14	Ryerson University	University	350 Victoria St.	CA	Toronto	Canada	at Gould St.	698	[350 Victoria St. (at Gould St.), Toronto ON M...	[{'label': 'display', 'lat': 43.65793504826254...	43.657935	-79.38049	NaN	M5B 2K3	ON	4ad7797cf964a520170b21e3
0	York University - Keele Campus	University	4700 Keele St.	CA	Toronto	Canada	NaN	16079	[4700 Keele St., Toronto ON M3J 1P3, Canada]	[{'label': 'display', 'lat': 43.77210701043441...	43.772107	-79.50219	NaN	M3J 1P3	ON	4ad78121f964a5208b0b21e3

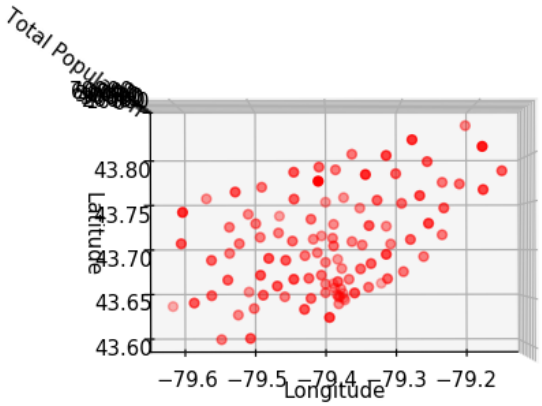
3. Exploratory Data Analysis

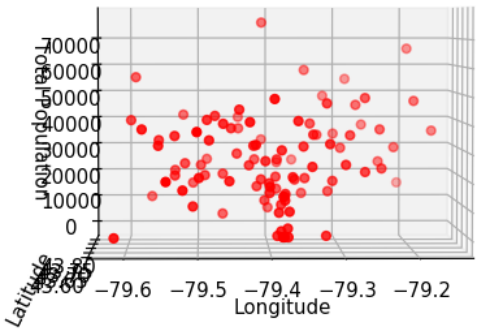
A choropleth map was created using the postal codes as boundaries. The scale runs from yellow to red, where yellow means less residents and red mean more residents. Universities were added to the map and are marked by a green marker.



A 3D scatter plot was then created to illustrate the distribution of population across each postal code area.



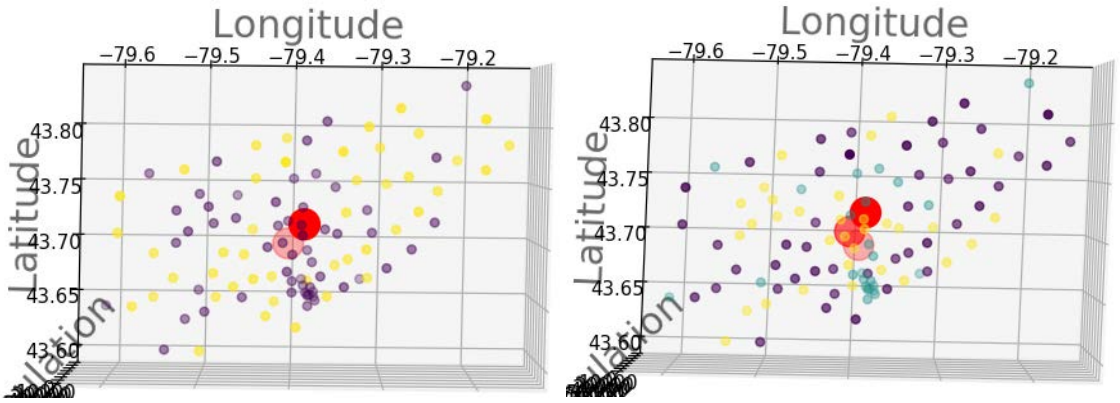




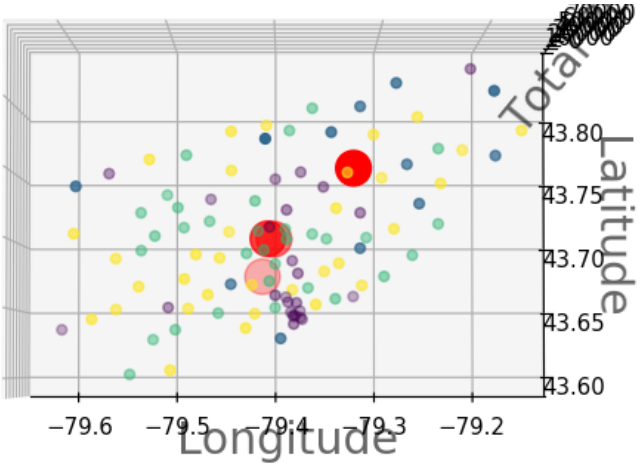
4. Predictive Modeling

K-means clustering was used to determine the ideal locations for universities in consideration to the population around each cluster.

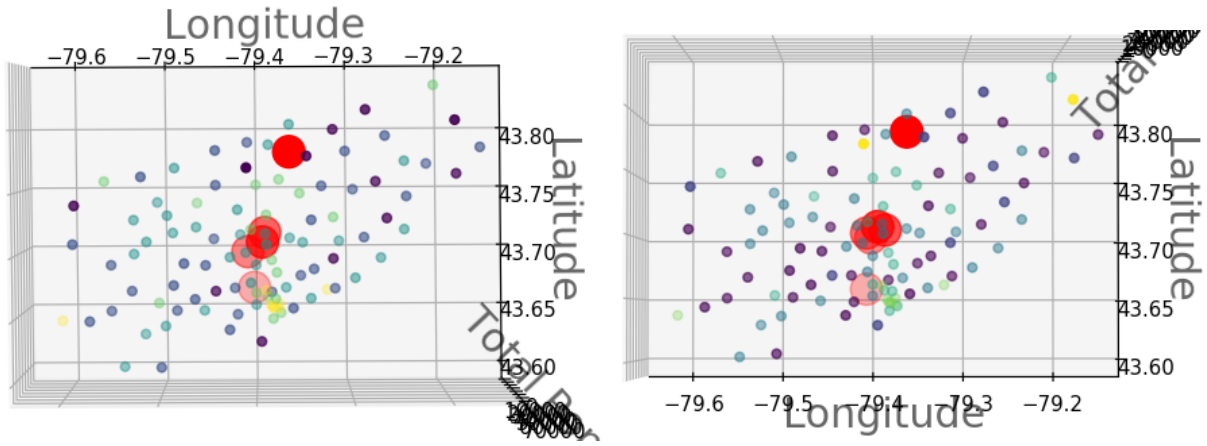
K-means = 2 and k-means = 3 showed results that were in similar positions to the University of Toronto and Ryerson University situated in Central Toronto.



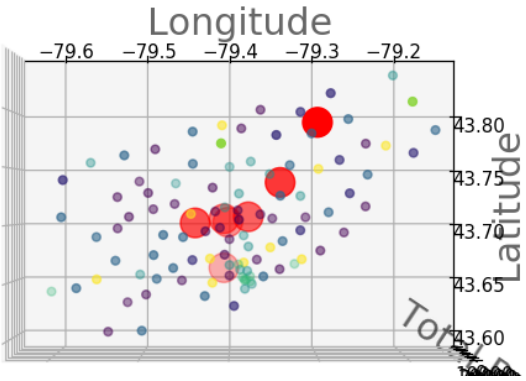
K-means = 4 showed an additional university plot at x = -79.3252, y = 43.7529 or roughly in the area of postal code M3A in the Parkwoods Neighborhood, North York.



K-means = 5 and k-means = 6 shows the additional university plot at x = -79.3645, y = 43.7820 or roughly in the area of postal code M2J in North York which consists of the neighborhoods of Fairview, Henry Farm, and Oriole.



K-means = 7 shows the furthest university plot from Central Toronto at x = -79.3014, y = 43.7884 or roughly in the postal code area of M1W in the Steeles West and L'Amoreaux West neighborhoods of Scarborough. Another data point of interest is located at x = -79.3305, y = 43.7451 which is around M3B in North Don Mills and near M3C in Don Mills South and Flemingdon Park neighborhoods, all of which are in North York.



Higher levels of k-means show k-means centroids that are in the general area of Central Toronto as well as Scarborough and North York.

5. Conclusions

With the assumption that Toronto does not have any universities, the best areas to place the first three universities are around the Central Toronto area due to the high density attributed to its numerous neighborhoods. A fourth university is ideally situated at the boundary of North

York (M2J) and Scarborough (M1W), and a fifth university is ideally located around the southeastern corner of North York (M3B and M3C).

With this, the University of Toronto and Ryerson University are situated at the ideal locations for universities in Toronto, while York University is not an ideal location since k-means never determined the area as an ideal location for the first 15 k-means.

Should the Ministry of Training, Colleges, and Universities decide to put up a public university or if any private universities decide to put an extra campus in Toronto, the most ideal location would be in postal code M2J in North York in any of the neighborhoods of Fairview, Henry Farm, or Oriole.

6. Future Directions

While the model only accounted for the city of Toronto, a more accurate model may be to include the other cities in the Greater Toronto Area, which include Durham, Halton, Hamilton, Peel, and York as well as other surrounding suburbs. This may be the reason why York University was not included in any of the k-means.

One other feature that could be included could be average income which could determine which type of school, public or private, is ideal for the centroids. The amount of commercial establishments could also be another factor as universities would typically have some commercial establishments nearby. Finally, the amount of available land in a location could determine the size and carrying capacity of the university and could therefore equate to a more accurate model.