Best Neighborhoods for Students in Toronto

By: Aref Jadda

Introduction

This project aims to find the best places in Toronto for a university student to live based on their personal preferences and lifestyle. This can help out students in their house search process especially if they are new to Toronto. To be more specific, we will only perform the analysis for students attending the University of Toronto, as it is the largest university in the city of Toronto. Students from universities close to University of Toronto, like Ryerson University in downtown Toronto will also be able to use this as a reference.

We will specifically only examine places in the city of Toronto, not the Greater Toronto Area (GTA) which includes neighboring towns and cities. The areas outside of the city will be too long of a commute for students and not very ideal.

We will use our knowledge of unsupervised machine learning to create different clusters based on criteria such as distance from the university, the number of grocery stores, and access to public transit.

Data Collection and Cleaning

Based on our goals for this project some of the data we will require are:

- The coordinates of different areas in Toronto based on their postal code
- The coordinates of the university
- The number of nearby grocery/convenience stores in each area
- The number of coffee shops near each area

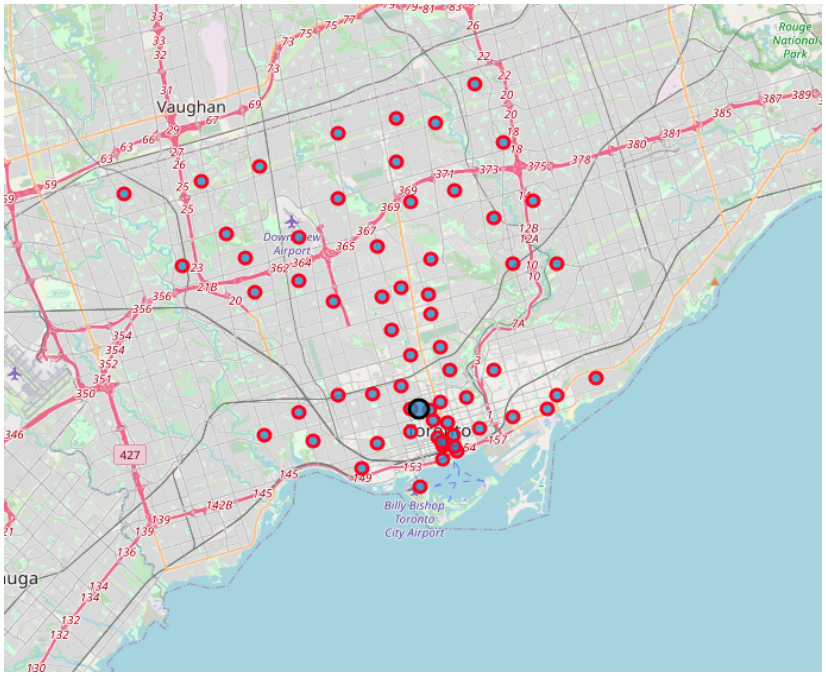
- The number of libraries near each area
- Whether there are public transit services nearby
- If the nearby transit is a subway station or bus station
- The number of gyms nearby

For collecting the data needed we have the four sources below:

1. Wikipedia for a table of Toronto areas by postal code
2. This link from IBM for getting the coordinates data for each postal code in Toronto
3. Google Maps for the coordinates of the University of Toronto
4. Foursquare API for gathering venue data for each postal code area

The data from scraped websites and collected through APIs from the sources above were initially very messy and tough to analyze. Therefore, we had to clean the collected data and put them into pandas dataframes readable by our machine learning algorithms.

After getting our final dataframe the postal code locations were visualized on a map using folium (the blue dot is the University of Toronto).



Methodology

This project was focused on the data for students attending the University of Toronto. Ergo, only data relevant to this university was collected for analysis. However, the exact same process can be performed for any other university around the world.

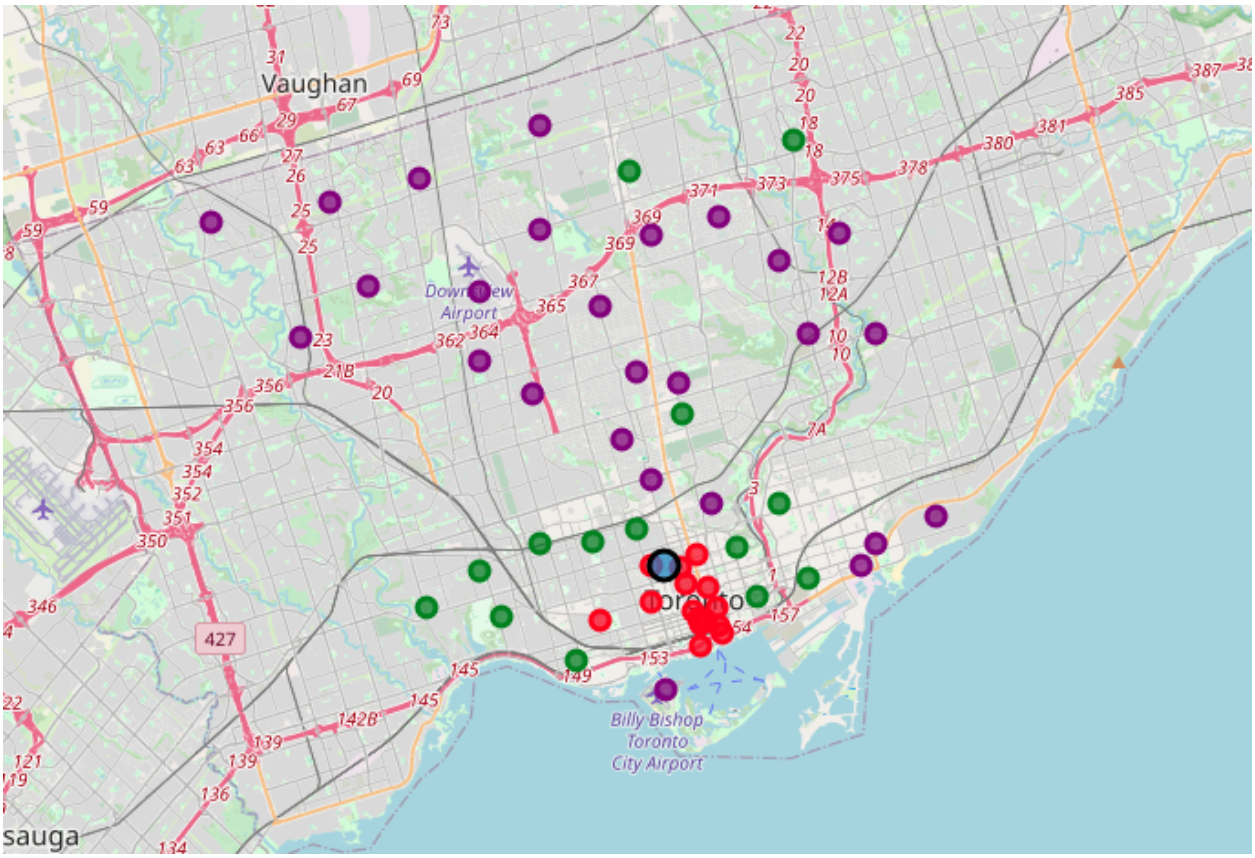
We started by collecting Toronto neighborhoods data from Wikipedia and then moved on to collecting each neighborhood's coordinates alongside the coordinates of the university. The coordinates would help us visualize the data on an interactive Folium map and we would also be able to use them to collect information about each area, specifically the venues close to them, for each single coordinate. We specifically chose a radius of 500 meters around each postal code, as that is a reasonably convenient walking distance.

Our next step was cleaning our gathered data into something that the machine can cluster, and that we would also be able to analyze easily after the clustering was done. We used only the categories of each venue as the details were not crucially important, and would just make the cleaning process longer. For analyzing each postal code, we then grouped the data by postal code and used the aggregate sum as we wanted to know how many of each general category of venue we had near each area.

In the final step we will be focusing on analyzing our clean data in order to differentiate between the different neighborhoods. We will use unsupervised learning, specifically k-means clustering, to cluster our data into areas that are similar to each other. We will most likely divide the neighborhoods into 3 or 4 clusters. After we will look at the properties that grouped the areas together and use a map to visualize everything.

Analysis

After clustering our data into 3 different clusters we now have the map below showing each cluster with their own unique color and the University of Toronto with a larger blue circle.



Let us now examine each cluster to see what patterns can be noticed among them.

Cluster 1:

This cluster consists of areas that are on average the furthest away from the university. As we can see many do not have a library or gym near them, some even not close to any convenience stores. These are the less crowded parts of the city in the central and northern parts of Toronto. Properties of some can be seen in the table below.

	Bus Station/Stop	Coffee Shop / Café	Grocery/Convenience/Drug Store	Gym	Library	Train/Subway Station	Distance to University
0	0	0	0	0	1	0	9.689
1	0	0	1	0	0	0	8.996
2	0	1	1	0	0	0	12.856
3	3	0	1	0	0	0	11.192
4	0	2	0	2	0	0	9.372
5	0	5	1	2	0	0	8.348
6	1	1	5	0	0	0	10.260

Cluster 2:

This is the area which represents the main downtown Toronto. All are very close to the university and have long lists of each category of venue close to them. Properties of some can be seen in the table below.

	Bus Station/Stop	Coffee Shop / Café	Grocery/Convenience/Drug Store	Gym	Library	Train/Subway Station	Distance to University
0	3	21	20	20	3	5	1.288
1	10	29	15	16	5	5	1.772
2	9	28	14	25	4	5	2.327
3	15	34	10	10	1	3	2.881
4	12	37	10	12	6	6	0.968
5	12	40	13	18	4	6	1.661
6	5	40	12	13	1	2	2.612
7	19	38	18	22	1	8	2.114
8	21	39	16	22	0	5	2.164
9	2	31	8	6	24	2	0.435
10	8	38	14	1	1	1	1.063
11	16	38	21	19	1	5	2.657
12	19	37	17	24	2	9	1.974

Cluster 3:

This cluster consists of mainly neighborhoods that are an average distance away from the university. They are mostly near subway stations or at the very least close to a bus stop. They almost all have many coffee shops, stores and gyms nearby. Some are close to a library but about half are not. Properties of these areas are seen in the table below.

	Bus Station/Stop	Coffee Shop / Café	Grocery/Convenience/Drug Store	Gym	Library	Train/Subway Station	Distance to University
0	10	9	5	2	1	3	12.563
1	1	11	9	2	0	1	10.798
2	7	11	8	4	1	1	4.659
3	2	11	4	5	0	1	5.488
4	1	9	2	8	1	0	4.200
5	3	12	8	3	0	0	2.848
6	8	15	7	9	0	1	3.611
7	1	10	5	6	0	1	1.399
8	1	9	3	0	0	0	2.767
9	5	8	7	3	2	0	4.696
10	6	14	12	4	3	1	4.165
11	6	10	8	2	1	0	6.908
12	5	8	5	0	0	0	6.221

Results and Discussion

The clustering separates the areas almost perfectly. We analyze the city of Toronto and managed to find three different categories of areas. Although this is a very basic model with a low number of features, it is quite descriptive and also very easy to analyze. Let us now discuss each area in more detail and determine pros and cons of each.

Cluster 1: Calmer Lifestyle, Longer Commutes

These are the areas in the quieter parts of the city without all the bright lights and the tall buildings stacked one after another. Some pros of these areas are that after long days at the university in downtown you can come home to a peaceful area away from all the noisy streets. You can also spend your leisure time in the beautiful parks and neighborhoods around, which are found all over Toronto. These areas are also on average on the cheaper side and you don't have to pay large sums of money for parking each month if you plan on owning a car.
The cons are that as expected there is not always a variety of each venue category located near you. Chances are that you will not be at walking distance to any grocery

stores, gyms, or cafes. Last but certainly not least, these are areas with the longest commute times to the university so they are not very ideal if you cannot stand spending around 1 to 2 hours each day in public transit to the university.

Cluster 2: Downtown Lifestyle, Neighboring the University

The second cluster is basically downtown Toronto in a nutshell. Toronto is amongst the busiest and most populated cities in North America. Thus, living there might not suit everybody's lifestyle preferences especially if you are used to smaller cities or towns. The streets are always noisy and filled with people rushing from building to building. Some may enjoy this lifestyle and get a boost of motivation from all the downtown energy and some may find it overwhelming.
Some amazing things about these areas are that you are very close to the university so your daily commute is very convenient. Being close to the university also gives you ease of access to all the universities amenities and facilities. You can use their amazing gyms and libraries with the comfort of being close to home. You are also in close proximity to all downtown cafes, restaurants, stores, etc. Living in these areas can get very expensive. Owning a car is almost never an option for a student living there. You might also have to sacrifice living in a much smaller space compared to areas in cluster 1 or 3.

Cluster 3: The Average

The last cluster can be seen as the average of the first two in almost every single aspect. They are in the middle in terms of distance to the university and commute times. They are certainly not as crowded as downtown areas and are busier than areas in cluster 1. They normally have a good number of each category of amenities close to them and they are almost all close to the metro station.
They are also average in terms of rental prices. You might be able to own a car and live in a decent sized space depending on the exact location, building, and of course your

own budget. These areas are the ones that most students prefer given that they are decently close to the university and still have all categories of venues near them.

Conclusion

This project aimed to find the best areas for university students to live given their preferences, and lifestyles. Many students come to Toronto from other places for studies and are not sure where to live. There are also students that have lived in Toronto previously that are not sure which areas would fit their needs the most. By collecting some information about the different areas in the city of Toronto that most students tend to care about when deciding on a location to live we tried to make this process of finding a home easier.

We created 3 different clusters of areas based on postal code and we analyzed all 3 to find the pros and cons of each. The first cluster was best for people that prefer quieter areas that do not have an issue with longer commutes. The second cluster was basically downtown Toronto near the university with all its glory and noisy streets. The third was an average of the first two, not too far from the university and not too close to the busy downtown areas.

The ultimate decision is for the students to make on which areas they prefer and what budgets they have. Based on the data I believe most people would prefer the third cluster, being in close proximity to different categories of venues and also decently close to downtown Toronto and the university. This analysis can be performed for any other city and around any other university to find the most ideal places for living.