



Best Neighborhoods for Students in Toronto

Applied Capstone Project
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
- We will only perform the analysis for students attending the **University of Toronto**
- We will specifically only examine places in the city of Toronto, not the Greater Toronto Area (GTA)
- 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**



The Problem

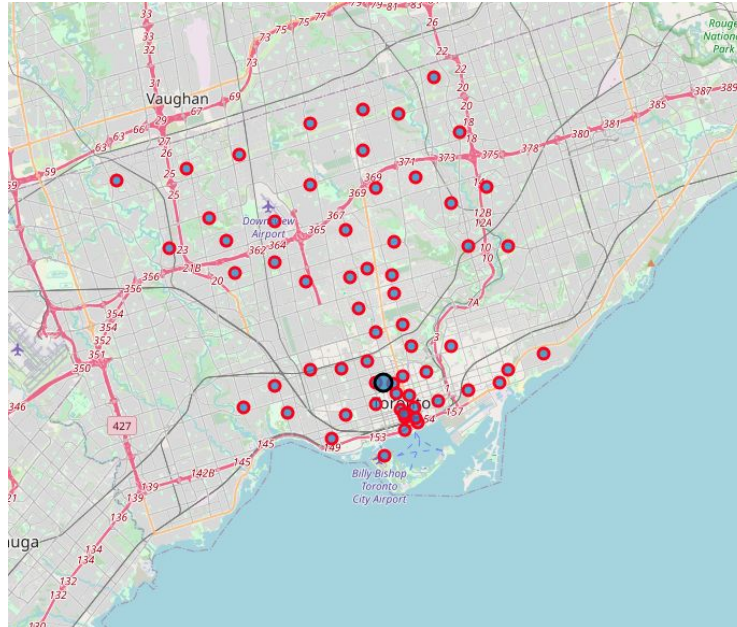
- How can we use data from the city of Toronto to find differences between different neighborhoods?
- Can we use this information to give reliable recommendations to students attending the University of Toronto?
- Some data we will require are:
 - ❑ The coordinates of different areas in Toronto based 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



Data Sources

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

Toronto Postal Codes Map





Methodology - Data Collection and Cleaning

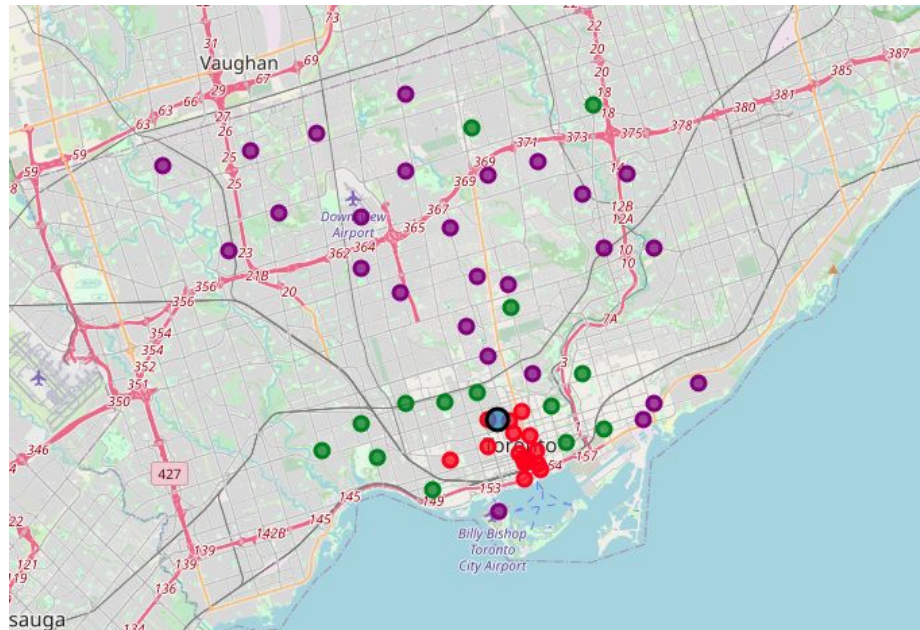
- Gather postal codes from Wikipedia and the respective coordinates of each
- Clean the gathered data into pandas dataframes
- Group entries by postal code
- Collect venue information from Foursquare for each postal code
- One-hot encode the Foursquare data and attach it to main dataframe
- Remove unwanted categories from the data and generalize the categories for ease of analysis



Methodology - Unsupervised ML

- Use the K-means library from sk-learn
- Try various different values of K
- Pick the best K (3 in this case)
- Visualize the clusters on a map and analyze their properties

Cluster Map





Cluster 1:

- Calmer lifestyle, longer commutes

	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
7	1	1		0	0	0	1	13.938
8	1	0		0	0	0	0	10.164
9	0	1		1	0	1	0	13.479
10	0	1		0	1	0	0	15.952
11	2	1		1	0	1	0	10.192
12	2	6		2	1	0	0	10.355



Cluster 2

- Downtown lifestyle, neighboring the university

	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

- The average cluster

	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



Conclusion

- This project aimed to find the best areas for university students to live given their preferences, and lifestyles
- 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 ultimate decision is for the students to make on which areas they prefer and what budgets they have