

IBM Data Science Capstone Final Project

Yijun L.

Feb. 2020

1. Introduction

In this project, I developed an application that would help students at the University of Toronto explore the cafes/restaurants near the school and the entire downtown Toronto area. As an international student at the University of Toronto myself, I had trouble finding good food around the campus, and I would have appreciated an application that lists a variety of options and creates a visual for me. My application also groups 37 neighborhoods in downtown Toronto into 5 areas and serves as a guide for students who are interested in exploring some of the city's most popular venues.

2. Data

I used location data from FourSquare, Toronto Postal Code, Borough and Neighborhood information from Wikipedia and geospatial data provided by Coursera.

In the first part of the project, I explored the cafes/restaurants near the University of Toronto. In the second part, I examined clusters within the

downtown Toronto areas and provide observations for each cluster. I also normalized the dataset to help interpret features with different magnitudes and distributions equally.

I only included venues in the downtown as it is where most of the venues worth visiting are concentrated in. Plus, most of students who are originally from outside Toronto live in the downtown area, so I only focused on the venues that are most relevant to them.

3. Methodology

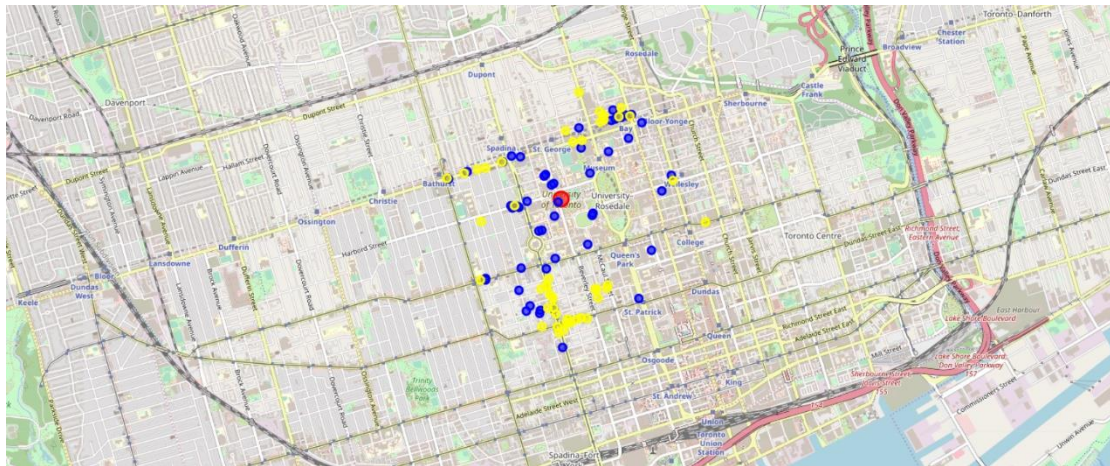
In the first part of the project, the final deliverable is a map centered around the University of Toronto, showing all the cafes/restaurants within a 1km radius. No machine learning methods are used here.

In the second part, I used K-means clustering to partition the 37 neighborhoods in downtown Toronto into 5 areas based on the frequency of occurrence of each venue category. The results were used to provide insights of what each area is known for, e.g. fancy restaurants, tourist attractions, etc.

The process is rather simple, so no exploratory data analysis or inferential statistical testing was required for this project.

4. Results and Discussions

The explorations into cafes and restaurants near the school gives us 50 cafes (marked as blue circles) and 50 restaurants (marked as yellow circles), as shown on the map below.



It shows that most diners are in these three areas (see the green circles), with the most concentrated in circle 3.



If a student is unsure about where to grab something to eat, she/he can always go to the circled areas on the map and find many options to choose from.

The exploration of different venues in Downtown Toronto indicates that 37 neighborhoods can be grouped based on their similarities quite reasonably. In cluster 0, we have 12 neighborhoods including Harbourfront East, Toronto Islands and Toronto Dominion Centre, etc. Some of the most common venues found within these neighborhoods are hotels, theaters, Italian restaurants and steakhouses. This area is, in fact, where most of the well-known attractions in Toronto are located. Many tourists would choose to stay and dine nearby, and thus the hotels and restaurants.

In cluster 2, which is comprised of 6 neighborhoods including Chinatown, University of Toronto and Kensington Market, etc, the most common venues are cafes, bars, vegetarian/vegan restaurants and a variety of Asian cuisines. A logical deduction would be this is the area where young people and students, especially students from the University of Toronto considering the size of the school, like to hang out. Also, Asian restaurants are relatively affordable options compared with steakhouses, and thus are perfect for the young people.

I also found similar patterns in cluster 3, which is a dynamic area known for good coffee shops, clothing stores and bars/pubs. It is another place where the students would enjoy going.

However, the most common venues shown in Cluster 0 are completely the same among the 7 neighborhoods. It is possibly because that these neighborhoods are relatively small in size, and so close to each other that they are not differentiated in the dataset.

Not surprisingly, Rosedale is indicated as an outlier. The venues are so different from other neighborhoods that it does not fall in any other clusters. In fact, Rosedale is home to some of Canada's richest and most famous citizens, and has been ranked the best neighborhood in Toronto to live in.

5. Conclusion

I analyzed cafes and restaurants options near the University of Toronto campus and a range of venues within the downtown area using Foursquare data and K-means clustering. The results are a fair representation of the reality and a potential city guide for college students who are new to the school and Toronto.

The scope of the study is limited to the venues in different neighborhoods, yet it can be taken further to include rental prices of each neighborhood, and thus provide useful information to students who try to decide the best

place to live in. Unfortunately, the only dataset I found that showed rental prices in Toronto went by completely different neighborhood names as given in Wikipedia. This task would be a lot less challenging for other cities where a standard list of neighborhood names is in place.