Finding out the similarity between Toronto’s neighborhood and Cherryhill

Jingwei Li

October 27, 2019

1. Introduction
   1. Background

Cherryhill is a community in London, Ontario. There are over ten apartment buildings in this area, making Cherryhill the most popular renting choice for Western University’s students. Most of those students will work in GTA area after they graduate and rent an apartment near their workplace. Some of them want to find out a place that is like to their old apartment.

* 1. Problem

This project aims to calculate the similarity between Cherryhill and Toronto’s different areas and find out the most similar area.

* 1. Interest

Cherryhill is the place where I currently live. However, after I graduate from university, I plan to find a job in Toronto. So, I am interested in this problem. Other students who facing the similar situation may also be interested.

1. Data acquisition and cleaning
   1. Data sourcesThe postcode, borough and neighborhood of Great Toronto’s Area can be found from Wikipedia (<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>). The latitude and the longitude of each postcode can be found from ([http://cocl.us/Geospatial\_data](https://cocl.us/Geospatial_data)). The postcode, borough, neighborhood, latitude and the longitude of Cherryhill can be find using google search. The venues around each postcode can be found by requiring from Foursquare.
   2. Data cleaning

Several steps are needed before we can get a useful dataframe.

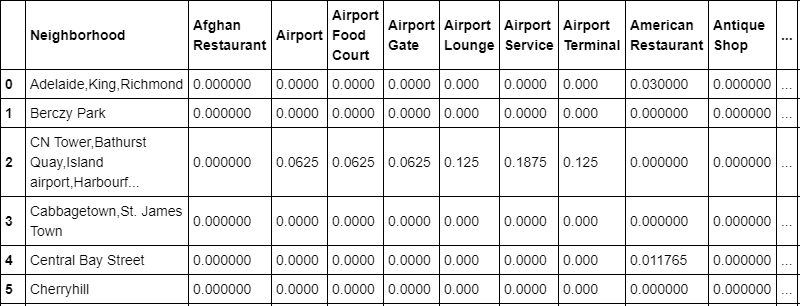
1. Extracting columns ‘PoatalCode’, ‘Borough’ and ‘Neighborhood’ from original dataframe.
2. Ignoring cells with a borough that is ‘Not assigned’.
3. Combining rows with same postcode.
4. Renaming the neighborhood’s cell as borough’s name if the cell was ‘Not assigned’.
5. Merging the dataframe with another dataframe which contains latitude and the longitude of each postcode.
6. Extracting rows only have ‘Downtown Toronto’ and merge with Cherryhill’s dataframe.

Table 1. Geomatic information of downtown Toronto and Cherryhill



1. Requiring venue information around each postcode from Foursquare.
2. Converting venue’s category into dummy variable
3. Grouping rows by neighborhood and taking the mean of the frequency of occurrence of each category

Table 2. Frequency of each category’s occurrence based on postal code



1. Exploratory Data Analysis
   1. Calculation of similarity

Euclidean distance is used to calculate the similarity between Toronto’s neighborhood and Cherryhill. The frequency of each category’s occurrence based on postal code is used as data. Finally, finding out the area with lowest Euclidean distance.

1. Result

Table 3. Euclidean distance between Toronto’s neighborhood and Cherryhill



From the chart, the neighborhood ‘Cabbagetown,St. James Town ' in downtown Toronto has the lowest Euclidean distance with Cherryhill. Therefore, the neighborhood ‘Cabbagetown,St. James Town ' is the most similar area to Cherryhill among downtown Toronto.

1. Discussion

Although the neighborhood ‘Cabbagetown,St. James Town ' is the most similar area to Cherryhill among downtown Toronto, the Euclidean distance is 0.570684 which is not small enough. Therefore, students who wants to find a similar living place in GTA should looking for other borough such as North York, Scarborough and Etobicoke.

1. Conclusion

People who wants to find out a similar living place in downtown Toronto to Cherryhill should choose neighborhood ‘Cabbagetown,St. James Town '.