Toronto Neighborhood Clustering based on Citizen Comfort Level

PRIAMBUDI PUJIHATMA

INTRODUCTION

- Purpose of this analysis is to cluster Toronto Area based on several factors that affect citizen comfortable living.
- The comfort level is defined by the following factors:
  - Apartment rental price
  - Presence of supporting facilities, such as: restaurant, coffee shop, park, health facilities / gym, grocery stores, farmers market, café and bakery.
- Output of this analysis is the neighborhood clusters along with plus delta aspect of each neighborhood.

DATA SOURCE

- List of Toronto borough, neighborhood and borough codes as shown in https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and coordinate for each borough
- Apartment rental price. The source for apartment rental price data is taken from Kaggle through this link: https://www.kaggle.com/rajacsp/toronto-apartment-price. The data is from 2018, which consists of 1125 apartment rental price data, along with the address, number of rooms, dens, bathrooms and coordinates for each apartment.
- The source of supporting facilities is taken from Foursquare. The default maximum number of events to be retrieved from Foursquare is 100. To achieve 100 venues per borough, the search radius is increased up to 5 km.

METHODOLOGY

APARTMENT DATA

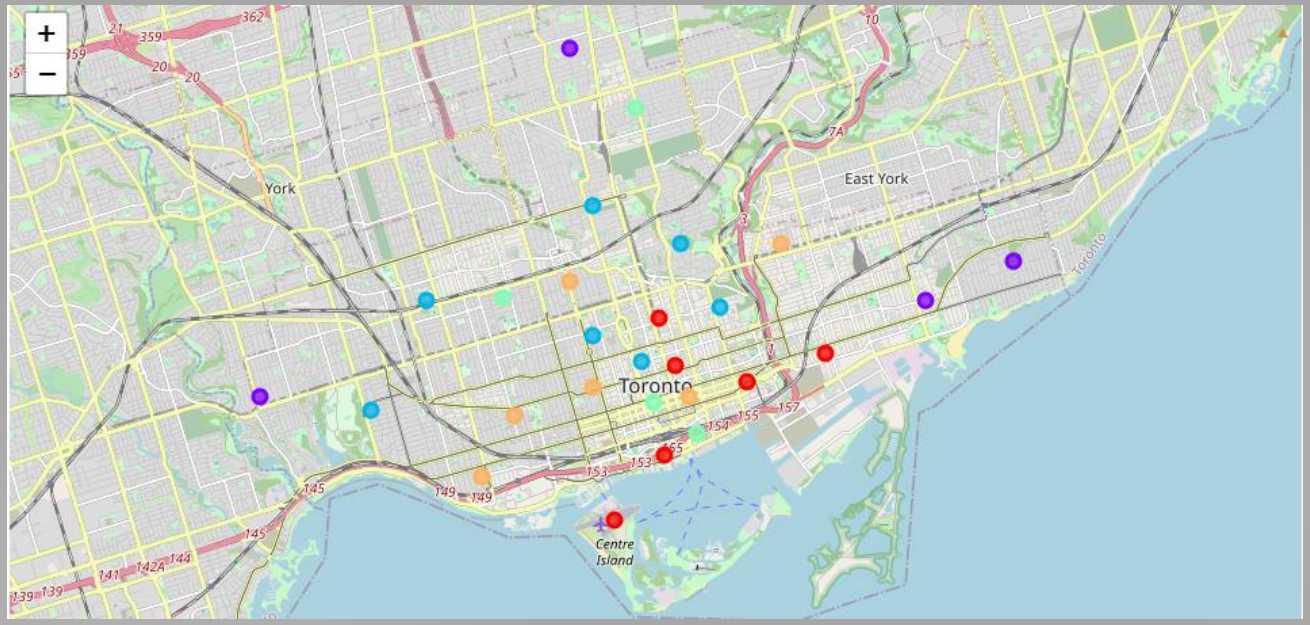
- Data preprocessing to get the neighborhood code from apartment rental data retrieved from Kaggle. The typical way to write a borough code on a Canadian address is by writing the borough code after the province, after the province code ON (Ontario). Thus, ON can be used as an identifier to split the address data and get the borough code value.
- Merge the apartment data using neighborhood as index
- Calculate the average rental data value for each neighborhood

METHODOLOGY

VENUE DATA

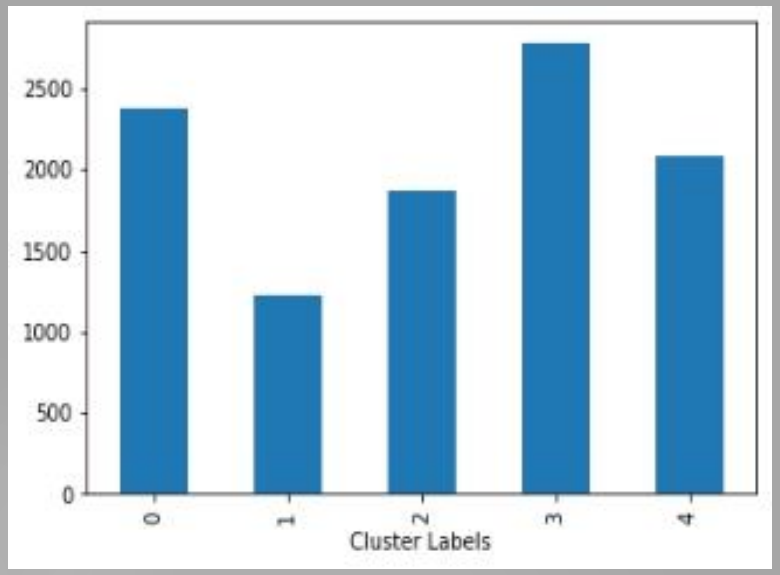
- The location data is taken using Foursquare API with 5 km search radius from each neighborhood center.
- The API resulted in 100 venues for each neighborhood, stored in the location data frame.
- The search result is then categorized based on venue type. Overall there are 165 unique venue categories.
- The average value of each category is calculated, which represent the ratio of each venue category when compared to overall venue available in each neighborhood.
- Out of 165 unique categories, 39 of them are different types of Restaurants,
- Thus, using Pandas, the columns that contains the word “Restaurant” in the column name are added together and grouped as one category “Restaurant”.

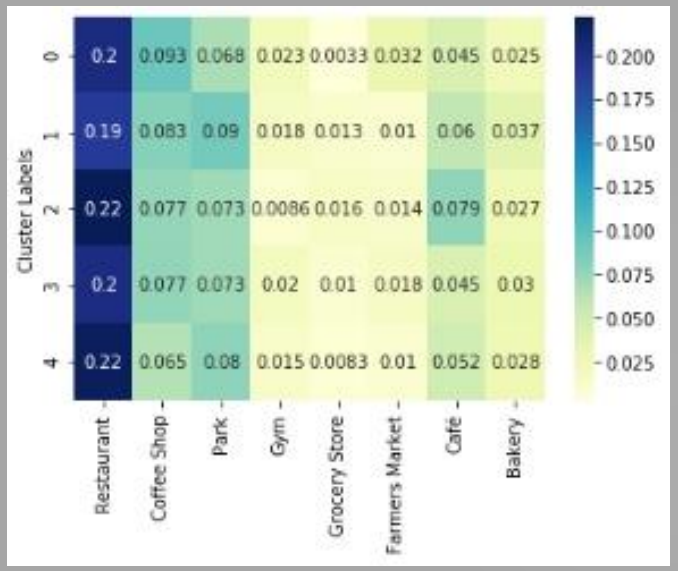
RESULT & DISCUSSION



Cluster Labels	Neighborhood
0	CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport, Church and Wellesley, Garden District, Ryerson, Harbourfront East, Union Station, Toronto Islands, Regent Park, Harbourfront, Studio District
1	India Bazaar, The Beaches West, North Toronto West, Lawrence Park, Runnymede, Swansea, The Beaches
2	Central Bay Street, Dufferin, Dovercourt Village, Parkdale, Roncesvalles, Rosedale, St. James Town, Cabbagetown, Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park, University of Toronto, Harbord
3	Berczy Park, Christie, Davisville, Richmond, Adelaide, King
4	Brockton, Parkdale Village, Exhibition Place, Kensington Market, Chinatown, Grange Park, Little Portugal, Trinity, St. James Town, The Annex, North Midtown, Yorkville, The Danforth West, Riverdale

RESULT & DISCUSSION





- Cluster 1 has the lowest average apartment cost, located at the edge of downtown Toronto. In terms of venue, Cluster 1 has more park and bakery compared to other clusters. However, Cluster 1 lacks of other supporting facilities, such as gym, grocery store, farmer marker and café.
- Cluster 0, 3 and 4 are “downtown clusters” that have high apartment cost with supporting facilities which outweigh cluster 1.
- Cluster 2 has medium average apartment cost. The location is still on the outer ring of cluster 0, 3 and 4 however they are located closer to downtown compared to Cluster 1. Cluster 2 can compete with Cluster 0, 3 and 4 in terms of supporting facilities.

CONLUSION

- This analysis provides Toronto area clustering based on comfort level, which is defined by apartment rental cost and the presence of supporting facilities.
- The results show that Cluster 1 is the area with the lowest rental cost, however it lacks the supporting facility when compared to other clusters.
- The best tradeoff is found in Cluster 2, with medium rental cost and supporting facility that can compete with downtown clusters.