Travel Guide for new visitor in Toronto

1. Introduction

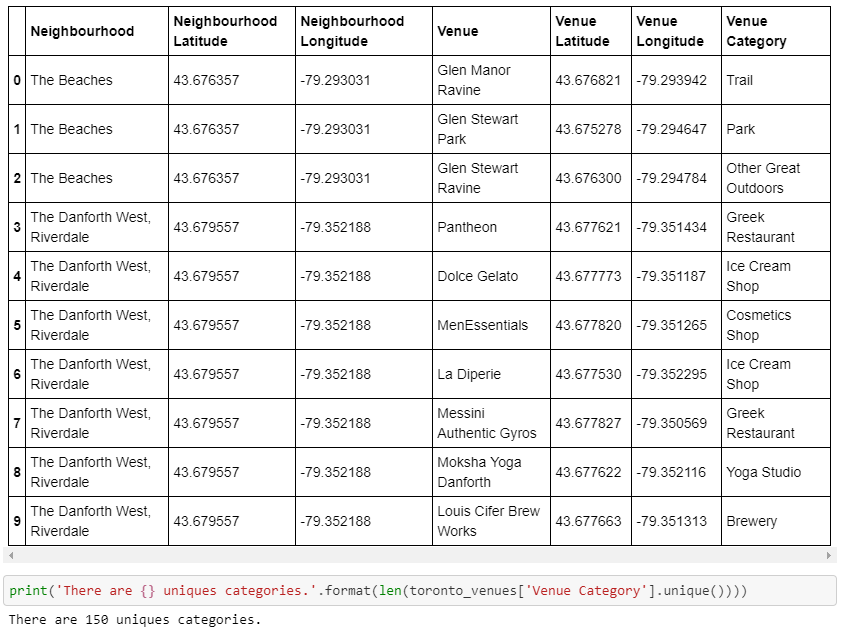
My report is focus on a travel guide for people going to visit Toronto but not familiar with there. My work is to simplify various venues and only list the most important and attractive venues. And this simplified neighborhood data will be a brief travel guide for new visitor in Toronto

1. Business Problem

Nowadays, people because of the more and more busy work, have restricted time to research a place and make a travel plan before vacation. In addition, as a new visitor it is very difficult to find out best places to have fun or best restaurant to have dinner. Besides, in order to travel on schedule, the transportation and hotel are also very important.

Previously, I retrieved all venues in East, West and Central Toronto by using Foursquare API as below:

(Venues data: <https://github.com/ysz951/Coursera_Capstone/blob/master/Toronto%20Venues.csv>)



We can see there are 150 categories contain Trail, Park, Other Great Outdoors and Ice Cream Shop, et al. So many different venue categories confuse people and make it hard to make a travel plan. Thus, simplifying the venue categories is necessary for new visitors.

The first step is ignoring the unnecessary venue, and then classify the rest categories into 6 groups: Entertainment, Food, Hotel, Scenic spot, Shop, Transportation. More details are described in Data Section.

1. Data

Pick up the useful data and classify them

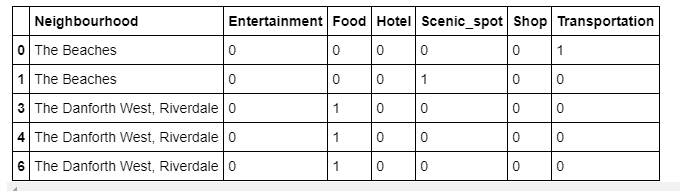


Venue before classified



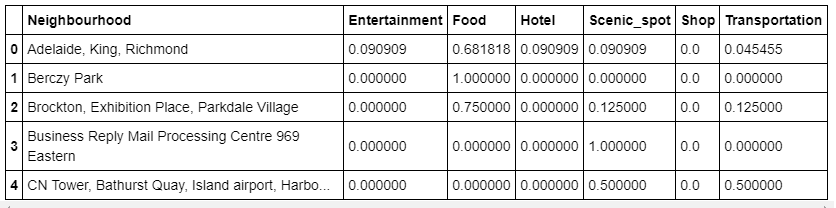
Venue after classified

And now the neighborhood data is very clear and succinct for new visitor

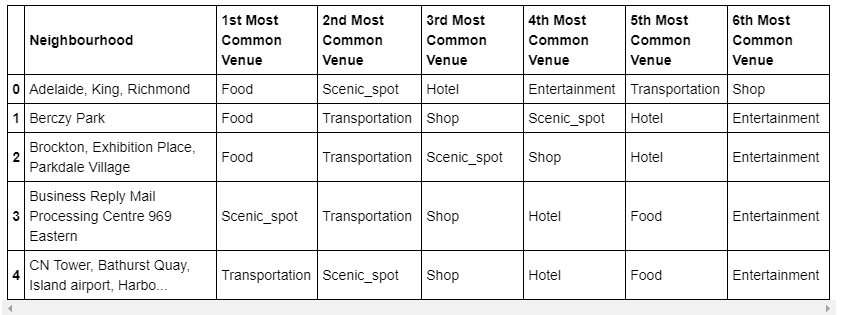


1. Methodology

Firstly, we group rows y neighbourhood and by taking the mean of the frequency of occurrence of each category

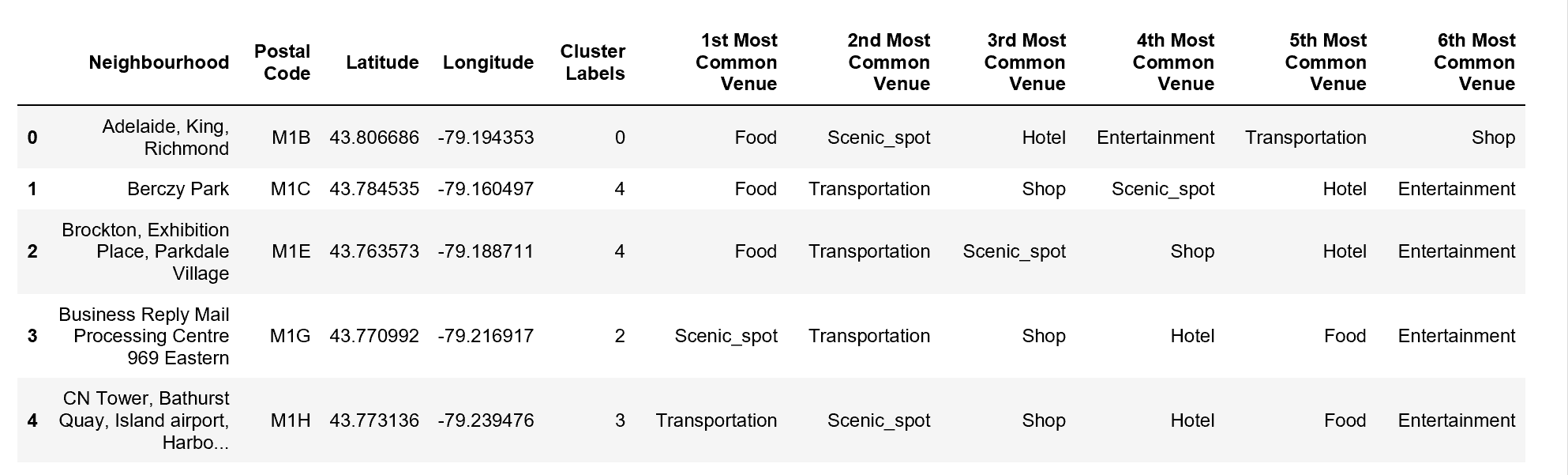


Secondly, we sort the venues in descending order



Finally, we cluster neighbourhoods and merge them with geospatial data

(<https://github.com/ysz951/Coursera_Capstone/blob/master/toronto_m.geospatial_data.csv>)



1. Result

I simplify the number of unique venue categories in Toronto from 150 to 93. Then I further classify the new venue categories into 6 types: Transportation, Scenic spot, Entertainment, Shop and Hotel. This simplified venue categories briefly list the most important and attractive venues for new visitor in Toronto. And neighborhoods are clustered and segmented by k means algorithm.

1. Discussion

From the result, new visitors can make their travel plan and choose which place to stay by their interests. If they prefer to take more time to enjoy the scenic spot in Toronto, they can choose to live in the hotel nearby scenic spot. If they can't wait to taste the delicious food, they can spend more time in the place surrounded by restaurants. Besides, for the people who change the schedule and need to leave next day, they can pick the hotel close to transportation to stay.

1. Conclusion

In order to make data clear and brief, we should drop some less important data and further classify them. And we can cluster and segment data by k means algorithm and get the information similar to what we most care or we want.