1 Data description and its usage

1.1 Data description

In order to cluster neighborhoods based on their popular venues, popular venues and their location-based data is required. In this project, we use the postcode as the basis to define a neighborhood. The data for the postcode and associated boroughs is obtained by scrapping relevant Wikipedia webpage (1). Data is cleaned and transformed to desired dataframes, as shown in Figure 1 below.

	Postcode	Borough	Neighbourhood
0	M1B	Scarborough	Rouge,Malvern
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union
2	M1E	Scarborough	Guildwood,Morningside,West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Figure 1: Some of the Postcodes, correspond boroughs and neigh hoods of Toronto

In next step geospatial, coordinates are assigned to these postcodes. These geospatial coordinates are used to obtain venues data from Foursquare (2). Venues are searched within the radius of 500 m form the geospatial location of these postcodes. Some of the venues associated with “The Beaches’ neighborhood are shown in Figure 2.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
1	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
2	The Beaches	43.676357	-79.293031	Starbucks	43.678798	-79.298045	Coffee Shop
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Beaches	43.676357	-79.293031	Dip 'n Sip	43.678897	-79.297745	Coffee Shop

Figure 2: Venues associated with 'The Beaches' neighborhood

Now we see how we use this data to solve our problem to find an optimal location for establishing the new restaurant.

1.2 Data usage to solve problem

Now we have the venue data for each neighborhood that we have obtained from Foursquare. We initially recognized that clustering will be most suitable option for this problem. As it will provide us cluster of neighborhoods with similar characteristics. Then we will be able to identify the most suitable neighborhoods for new investment opportunity. The steps involved in data usage to solve the process are summarized in the in Figure 3.

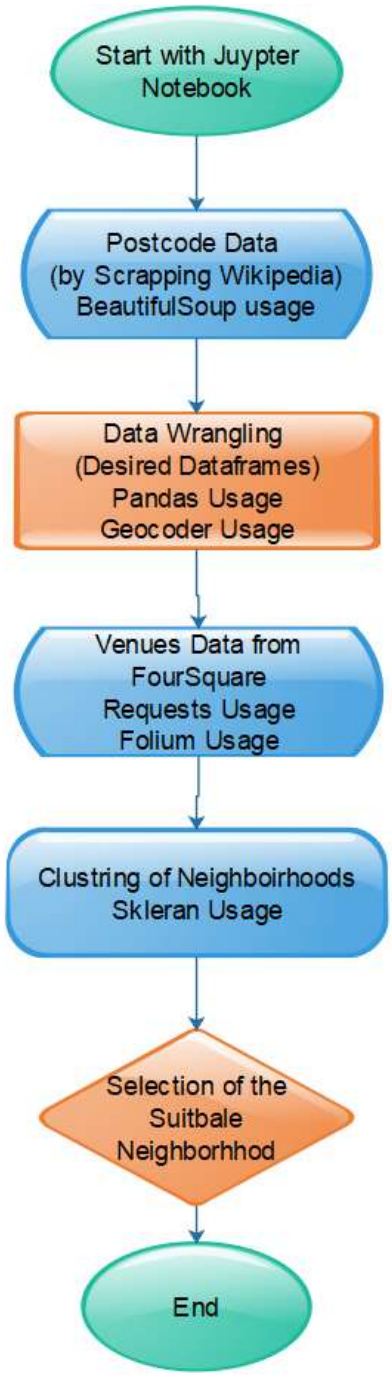

Figure 3: High-level process flow diagram of data usage

2 References

1. Wikipedia. List of postal codes of Canada 2019 [Available from: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
2. FourSquare. Places by Foursquare 2019 [Available from: https://foursquare.com/.