Battle of the Neighborhoods in Toronto

IBM® Applied Data Sciences Capstone Project

Introduction

- This capstone project utilizes the Foursquare data to study the market for restaurants and service oriented small businesses in the Downtown Toronto area.
- For the analyses of this study, K-means clustering is used to arrange the neighborhoods and venues into clusters.
- The study explores the neighborhoods that currently have a higher concentration of restaurant businesses, and areas where there are opportunities for opening service oriented businesses.

Problem Description and Background

- In week 3 we learned in this class, how to use and access Foursquare data.
- As part of the week 3 assignment, we also segmented and clustered Toronto neighborhoods.
    - So, it was determined that applying the foursquare data to compare the potential strengths, opportunities, and competition for opening new restaurants or small service oriented businesses in Toronto will be an interesting study.
- Therefore, I decided to pursue my analysis to examine the potential for opening restaurants and service oriented businesses by comparing different Toronto neighborhoods.

Data Description

- **The data for this study was obtained through two different sources:**
  - Postal codes and neighborhood information for Toronto was obtained from the following wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). The data for Toronto was then scraped and cleaned as described below and a data frame include postal code, neighborhood and borough information was then constructed.
  - Information on venues located in different neighborhoods were obtained from the Foursquare. As suggested in week 3, a free developer account was first setup with foursquare.com. Client Id, client secret information were obtained.
  - Finally, using a version number, the data for downtown Toronto was then extracted and merged with the table created above. These processes are explained below.

Data Scraping and Data Cleaning

- The postal code, neighborhood, and boroughs related information were cleaned and duplicates were integrated to construct a new dataframe.

]:

	PostalCode	Borough	Neighborhood
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

Next the latitude and longitude information were Obtained based on the postal codes

	PostalCode	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Next the Foursquare data was merged with the locational table for Toronto Downtown

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Adelaide, King, Richmond	100	100	100	100	100	100
Berczy Park	57	57	57	57	57	57
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	12	12	12	12	12	12
Cabbagetown, St. James Town	43	43	43	43	43	43
Central Bay Street	82	82	82	82	82	82
Chinatown, Grange Park, Kensington Market	92	92	92	92	92	92
Christie	17	17	17	17	17	17
Church and Wellesley	85	85	85	85	85	85
Commerce Court, Victoria Hotel	100	100	100	100	100	100
Design Exchange, Toronto Dominion Centre	100	100	100	100	100	100
First Canadian Place, Underground city	100	100	100	100	100	100
Harbord, University of Toronto	36	36	36	36	36	36
Harbourfront	48	48	48	48	48	48
Harbourfront East, Toronto Islands, Union Station	100	100	100	100	100	100
Queen's Park	41	41	41	41	41	41
Rosedale	4	4	4	4	4	4
Ryerson, Garden District	100	100	100	100	100	100
St. James Town	100	100	100	100	100	100
Stn A PO Boxes 25 The Esplanade	97	97	97	97	97	97

Next 'one hot encoding' is applied to obtain the different venue categories

	Neighborhoods	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Baby Store	Bagel Shop	Bakery
0	Rosedale	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Rosedale	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Rosedale	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Rosedale	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Cabbagetown, St. James Town	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

K-Mean Clustering is then run to sort the neighborhoods into 5 clusters

Exploration of Cluster 1

In [139]:
```
dt_merged.loc[dt_merged['Cluster Labels'] == 0, dt_merged.columns[[2] + list(range(5, dt_merged.shape[1]))]]
```

Out[139]:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Cabbagetown, St. James Town	0	Café	Coffee Shop	Pizza Place	Restaurant	Pub	Bakery	Italian Restaurant	Liquor Store	Pet Store	Pharmacy
2	Church and Wellesley	0	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Gay Bar	Restaurant	Burger Joint	Gym	Bubble Tea Shop	Hotel	Yoga Studio
4	Ryerson, Garden District	0	Coffee Shop	Clothing Store	Café	Cosmetics Shop	Bakery	Middle Eastern Restaurant	Theater	Sporting Goods Shop	Bubble Tea Shop	Restaurant
5	St. James Town	0	Café	Coffee Shop	Restaurant	Hotel	Clothing Store	Cosmetics Shop	Beer Bar	Cocktail Bar	Breakfast Spot	Italian Restaurant
6	Berczy Park	0	Coffee Shop	Cocktail Bar	Farmers Market	Seafood Restaurant	Steakhouse	Bakery	Beer Bar	Cheese Shop	Café	Diner
8	Adelaide, King, Richmond	0	Coffee Shop	Café	Steakhouse	Bar	Restaurant	Burger Joint	Sushi Restaurant	Asian Restaurant	Thai Restaurant	Gastropub
9	Harbourfront East, Toronto Islands, Union Station	0	Coffee Shop	Aquarium	Italian Restaurant	Hotel	Café	Scenic Lookout	Restaurant	Brewery	Fried Chicken Joint	Pizza Place

Exploration of Cluster 2

Exploration of Cluster 2:

In [140]:
```
dt_merged.loc[dt_merged['Cluster Labels'] == 1, dt_merged.columns[[2] + list(range(5, dt_merged.shape[1]))]]
```

Out[140]:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Rosedale	1	Park	Playground	Trail	Dessert Shop	Ethiopian Restaurant	Empanada Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop

Exploration of Cluster 3

Exploration of Cluster 3

[141]:
```
dt_merged.loc[dt_merged['Cluster Labels'] == 2, dt_merged.columns[[2] + list(range(5, dt_merged.shape[1]))]]
```

t[141]:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
14	CN Tower, Bathurst Quay, Island airport, Harbo...	2	Airport Lounge	Airport Service	Harbor / Marina	Sculpture Garden	Airport Food Court	Airport Terminal	Boat or Ferry	Boutique	Rental Car Location	Airport

Exploration of Cluster 4

Exploration of Cluster 4

In [142]:
```
dt_merged.loc[dt_merged['Cluster Labels'] == 3, dt_merged.columns[[2] + list(range(5, dt_merged.shape[1]))]]
```

Out[142]:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Harbourfront	3	Coffee Shop	Park	Bakery	Pub	Café	Mexican Restaurant	Breakfast Spot	Restaurant	Farmers Market	Spa
7	Central Bay Street	3	Coffee Shop	Café	Italian Restaurant	Burger Joint	Sandwich Place	Ice Cream Shop	Chinese Restaurant	Japanese Restaurant	Bubble Tea Shop	Bar
18	Queen's Park	3	Coffee Shop	Gym	Park	College Auditorium	Smoothie Shop	Sandwich Place	Burger Joint	Burrito Place	Café	Portuguese Restaurant

Exploration 5

Exploration of Cluster 5

In [143]:
```
dt_merged.loc[dt_merged['Cluster Labels'] == 4, dt_merged.columns[[2] + list(range(5, dt_merged.shape[1]))]]
```

Out[143]:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
17	Christie	4	Grocery Store	Café	Park	Candy Store	Diner	Italian Restaurant	Baby Store	Athletics & Sports	Restaurant	Coffee Shop

Results

- The results of this study are shown above. The results indicate that when we use the K-means clustering for the Toronto downtown area, if arranges the venues across 5 different clusters.
- Cluster 1 has a heavier concentration of restaurants and coffee shops. Clusters 2 and 5 appear more residential with a higher concentration of grocery stores, parks, playgrounds candy stores, baby stores etc.
- Cluster 3 is the area around the airport. Finally, cluster 4 also has a high concentration of restaurants, coffee shops and other food places. Cluster 4 appears to be an area that caters to students and young adults.

Discussions and Conclusion

- The results from this study indicate that while Clusters 1 and 4 present the biggest market for opening restaurants, there is also a lot of competitions given the heavy concentration food businesses in these areas.
  - Perhaps opening a restaurant in the Cluster 2 or Cluster 5 will have higher risk but also the potential for greater opportunities. Cluster 3, which is the airport area is also another possibility, since Toronto has a large international airport and many tourists and travelers pass by this area everyday.
- For small service oriented businesses such as laundry, day care, plumbing, electrical, financial services etc. the residential clusters 2 and 5 have the most opportunity. Additionally, Cluster 3 that has educational institutions nearby, also provides the opportunity for opening of some of these businesses as well.
- Overall, the capstone case was beneficial exercise, which helped me pull all of the learning from the previous courses and integrating the tools in analyzing this case study. I felt like I learned a lot about geospatial analysis, and obtaining Restful API from a service like Foursquare.