Battle of Neighborhoods with Universities in Toronto

Weiran Zhao

Feb 1, 2021

1. Introduction

Toronto is a metropolis with lots of universities and about 87,083 student population, which is a large number of potential customers. By comparing the neighborhoods with different universities in Toronto city, we will have a view of the similarities and differences of venues categories in the neighborhoods. Therefore, businessmen can have a better view of the surrounding area, make data driven discussions, chose location, or change business strategy.

2. Data acquisition and cleaning

2.1 Data sources

Toronto neighborhood data is taken from a Wikipedia page, that has all the information to explore and cluster the neighborhoods in Toronto. By finding the matched postal code of university, we can get the corresponding neighborhoods names.

Foursquare API is being used to get venues details for neighborhoods in Toronto. We will use explore query, search query and other ways to fully utilize the API functionalities. It provides the universities coordinates and information about categories of venues in neighborhoods.

2.2 Data Cleaning

Postal code of Toronto table is gathered from the Wikipedia website, which is from a HTML format page. Therefore, we need to clean up other tags and keep the table only, and transform it into data frame format. Then, we need to clean up the rows that with “not assigned” value from Borough column and combined the Neighborhoods with the same Postal Code.

```
# getting data
url='https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
raw_data= requests.get(url).text
#parse HTML codes
soup_data = bs(raw_data, 'html.parser')
#print(soup_data.prettify())
header = soup_data.find_all("table")[0].find("tr")
print(header)
```

	Postal Code	Borough	Neighbourhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

From the Foursquare API, we can easily get the addresses of the universities that we want to research. By matching the same postal code from the previous table, we can get the neighborhoods where universities allocated.

```
address = 'University of Toronto'
geolocator = Nominatim(user_agent="foursquare_agent")
location = geolocator.geocode(address)
latitude_uoT = location.latitude
longitude_uoT = location.longitude
postcode_uoT = location.address
print(latitude_uoT, longitude_uoT, postcode_uoT)
```

```
43.663461999999996 -79.39775965337452 University of Toronto, Wellesley-Hoskin Cycle Track, Bloor Street Culture Corridor, University–Rosedale, Old Toronto, Toronto, Golden Horseshoe, Ontario, M5S 1C4, Canada
```

	index	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	58	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568
1	66	M5S	Downtown Toronto	University of Toronto, Harbord	43.662696	-79.400049
2	85	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

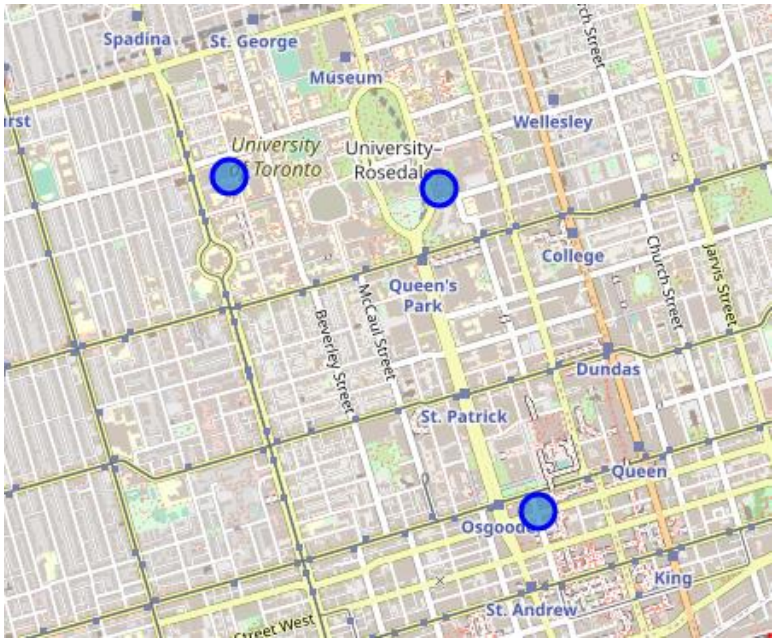
2.3 Feature selection

The venues categories are selected as the features for analysis, including Arts & Entertainment, Event, Food, Nightlife Spot, Outdoors & Recreation, Professional & Other Places, Residence, Shop & Service, Travel & Transport. Using the number of these features can find out what category of venues are taking most portion of the university neighborhoods.

	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport	University
0	46	86	9	152	59	73	138	66	110	100	OCAD University
1	31	107	7	119	53	40	140	69	98	51	University of Toronto
2	27	109	6	81	24	52	144	69	103	58	York University

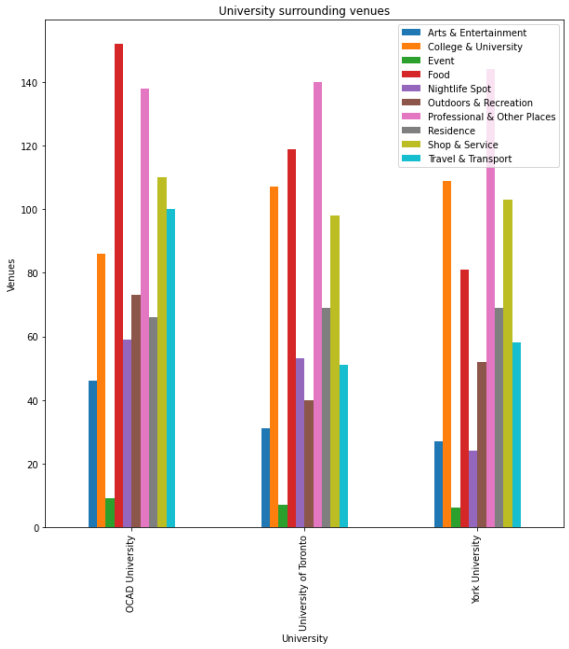
3. Exploratory Data Analysis

Three universities are chosen in this research – OCAD University, York University and University of Toronto. From the generated Toronto map, we can easily see the three university neighborhoods are allocated in the Toronto downtown area, but still have some distance. Therefore, it will be easy to compare the differences and find out the shortage service category since the they are located in the same area.

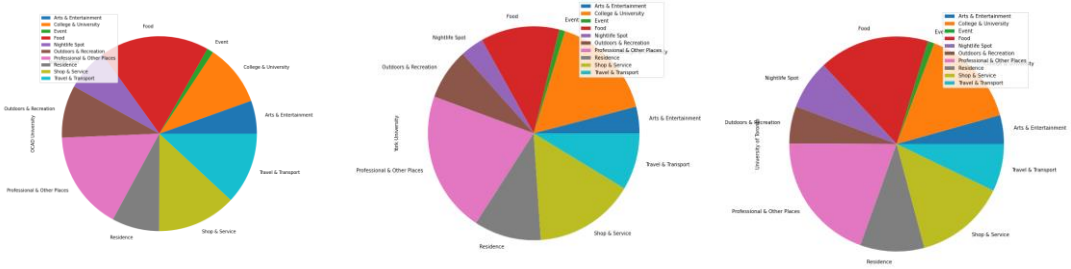


The bar chart shows the number of venues in different universities area. From the chart, we can see the neighborhoods' constitution are similar, which have almost same number of Professional & Other places, Shop & Service, Residence, and a small number of Event venues.

However, compared with other two universities, York University have a smaller number of Food and Nightlife Spots.



We can get the same conclusion from the pie charts. The universities surrounding businesses structures are similar, that all three neighborhoods have full facilities. The differences are on the Food and Nightlife Spot fields.



4. Conclusion

From the above data and graphs, we can see the neighborhood around York University will have less rivals on food and nightlife businesses, which is a potential business development area. Besides, from the bar chart, we can see the number of Professional & other places and Residence are higher than other two area, that means there exists enough potential clients in this area. The constant number of transportation facilities is also a guarantee for the customer flow volume. Therefore, have a business on food or nightlife fields at York University neighborhood is suggested.