Capstone Project--Solving Problems of University Students

Huimin Jiang

Dec 30, 2020

1. Description

1.1 *Determine the Problem*

For this assignment, we are required to define our own questions using the knowledge of API and location data to analyze the realistic problems. The ploblem I choose is how to choose a good restaurant and a place for fun for a student in University of Toronto to spend weekend time. It is because I am a university student now that this problem is highly correlated with my daily life, so solving this problems may be realistic to the sutdents like me. Moreover, as we clustered Toronto in previous weeks, I took university of Toronto as an example.

1.2 *Describe the Data*

The data required for this assignment can be acquired by Foursquare API--the geographical location and the ratings of restaurants. It is necessary to cluster Totonto and find the best place for students to go in Toronto. Wikipedia can provide the postal codes, https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and the geographical coordinates--latitutudes and longitudes of each neighborhood--can be acquired from the link, http://cocl.us/Geospatial_data. The data should be cleaned and modified and then construct a dataframe to analyze the problem.

2. Evaluation of Restaurants

2.1 *Load and analyze the data*

Before we load and prepare the data, it is necessary to import the libraries we need. Libraries for displaying images and plotting images are required. Then we can get the geographical coordinates of University of Toronto by geocode, and the coordinates are 43.663461999999996, -79.39775965337452. After defining the URL and sending the requests to Foursquare API, we can get the venues near university of Toronto. Since we are going to solve the daily problems of university students, we first evaluate the restaurants around the university by Foursquare API.

By sending the requests to Foursquare API, we can get full data of the venues around university of Toronto. Then we can modify the data into a data frame (figure 1) by extracting the names, categories and so on of the restaurants.

	name	categories	address	crossStreet	lat	lng	labeledLatLngs	distance	postalCode	cc	neighborhood	city	state
0	Punto Gelato, Simply Italian	Ice Cream Shop	146 Cumberland St	btwn Avenue Rd & Bay St	43.669955	-79.392603	[{'label': 'display', 'lat': 43.66995452843031...	833	M5R 1A8	CA	Yorkville	Toronto	ON
1	Scaddabush Italian Kitchen & Bar	Italian Restaurant	382 Yonge Street, Unit #7	Gerrard	43.658920	-79.382891	[{'label': 'display', 'lat': 43.65892029202872...	1299	M5B 1S8	CA	NaN	Toronto	ON
2	Fabbrica Rustic Italian	Italian Restaurant	66 Wellington St W	NaN	43.647161	-79.381691	[{'label': 'display', 'lat': 43.647161, 'lng':...	2228	M5K 1E7	CA	NaN	Toronto	ON
3	The Italian Cultural Institute	Art Gallery	496 Huron St	NaN	43.668432	-79.402169	[{'label': 'display', 'lat': 43.66843245214709...	657	NaN	CA	NaN	Toronto	ON
4	Papa Ceo Fine Italian Foods & Pizza	Pizza Place	654 Spadina Avenue	NaN	43.663300	-79.402317	[{'label': 'display', 'lat': 43.66329982959619...	367	M5S 2H7	CA	NaN	Toronto	ON
5	Italian Consulate Toronto	Embassy / Consulate	136 Beverley St	Dundas Street	43.654027	-79.394104	[{'label': 'display', 'lat': 43.65402694219784...	1090	NaN	CA	NaN	Toronto	ON

Figure 1

2.2 *Visualization*

By using the latitude and longitude of each restaurant in Figure 1, we can display the map of the restaurants around university of Toronto as follow (Figure 2):

Figure 2

2.3 *Further analysis*

From Figure 1, we can get the id of each restaurants as Figure 3 shows at tails.

egories	address	crossStreet	lat	lng	labeledLatLngs	distance	postalCode	cc	neighborhood	city	state	country	formattedAddress	id
Cream Shop	146 Cumberland St	btwn Avenue Rd & Bay St	43.669955	-79.392603	[{'label': 'display', 'lat': 43.66995452843031...	833	M5R 1A8	CA	Yorkville	Toronto	ON	Canada	[146 Cumberland St (btwn Avenue Rd & Bay St), ...	4e31afdd091a973ec9c5a2b5
Italian staurant	382 Yonge Street, Unit #7	Gerrard	43.658920	-79.382891	[{'label': 'display', 'lat': 43.65892029202872...	1299	M5B 1S8	CA	NaN	Toronto	ON	Canada	[382 Yonge Street, Unit #7 (Gerrard), Toronto ...	52f6816f11d24a43115dc834
Italian staurant	66 Wellington St W	NaN	43.647161	-79.381691	[{'label': 'display', 'lat': 43.647161, 'lng':...	2228	M5K 1E7	CA	NaN	Toronto	ON	Canada	[66 Wellington St W, Toronto ON M5K 1E7, Canada]	5b897e92db1d81002c91df8c
Gallery	496 Huron St	NaN	43.668432	-79.402169	[{'label': 'display', 'lat': 43.66843245214709...	657	NaN	CA	NaN	Toronto	ON	Canada	[496 Huron St, Toronto ON, Canada]	4db9bb6db5928d7fda728fc4
Pizza Place	654 Spadina	NaN	43.663300	-79.402317	[{'label': 'display', 'lat': 43.66329982959619...	367	M5S 2H7	CA	NaN	Toronto	ON	Canada	[654 Spadina Avenue, Toronto ON M5S 2H7,	4ad4c05ff964a52014f720e3

Figure 3

By inputting the id and defining the URL, we can get the ratings of each restaurant. The ratings of the three nearest Italian restaurants are as follows (Figure 4, 5 and 6):

```
venue_id = '5b897e92db1d81002c91df8c' # ID of Fabbrica Rustic Italian
url = 'https://api.foursquare.com/v2/venues/{}?client_id={}&client_secret={}&oauth_token={}&v={}'.format(venue_id, CLIENT_ID, CLIENT_SECRET,ACC

result = requests.get(url).json()
try:
    print(result['response']['venue']['rating'])
except:
    print('This venue has not been rated yet.')
```

```
This venue has not been rated yet.
```

Figure 4

```
venue_id = '52f6816f11d24a43115dc834' # ID of Scaddabush Italian Kitchen & Bar
url = 'https://api.foursquare.com/v2/venues/{}?client_id={}&client_secret={}&oauth_token={}&v={}'.format(venue_id, CLIENT_ID, CLIENT_SECRET,ACC
url
try:
    print(result['response']['venue']['rating'])
except:
    print('This venue has not been rated yet.')
```

```
5.2
```

Figure 5

```
venue_id = '52f6816f11d24a43115dc834' # ID of Scaddabush Italian Kitchen & Bar
url = 'https://api.foursquare.com/v2/venues/{}?client_id={}&client_secret={}&oauth_token={}&v={}'.format(venue_id, CLIENT_ID, CLIENT_SECRET,ACC
url
result = requests.get(url).json()
print(result['response']['venue'].keys())
result['response']['venue']
try:
    print(result['response']['venue']['rating'])
except:
    print('This venue has not been rated yet.')
```

```
dict_keys(['id', 'name', 'contact', 'location', 'canonicalUrl', 'categories', 'verified', 'stats', 'url', 'price', 'hasMenu', 'likes', 'lik
e', 'dislike', 'ok', 'rating', 'ratingColor', 'ratingSignals', 'menu', 'allowMenuUrlEdit', 'beenHere', 'specials', 'photos', 'reasons', 'de
scription', 'page', 'hereNow', 'createdAt', 'tips', 'shortUrl', 'timeZone', 'listed', 'hours', 'popular', 'seasonalHours', 'defaultHours',
'pageUpdates', 'inbox', 'attributes', 'bestPhoto', 'colors'])
7.5
```

Figure 6

We can see from the figures above that the rating of Scaddabush Italian Kitchen & Bar is the highest, 7.5. We can explore the restaurants further by finding the tips of different customers. We can only see 2 of the tips since the account is personal.

3. Evaluation of Places in Toronto

3.1 *Load and prepare the data*

After choosing the restaurant-- Scaddabush Italian Kitchen & Bar, we can find the proper area to stay for students in University of Toronto. We first download all the dependencies and then load the data we need from Wikipedia https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and the geographical coordinates--latitudes and longitudes of each neighborhood--from the link, http://cocl.us/Geospatial_data. The data we need are the postal codes, boroughs and neighborhoods from Wikipedia; the data we need are latitudes and longitudes from geospatial data. Then we can build a data frame as follow with these data.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

3.2 *Clustering Toronto areas*

First, we can build a map of Toronto with these postal codes and geographical coordinates.

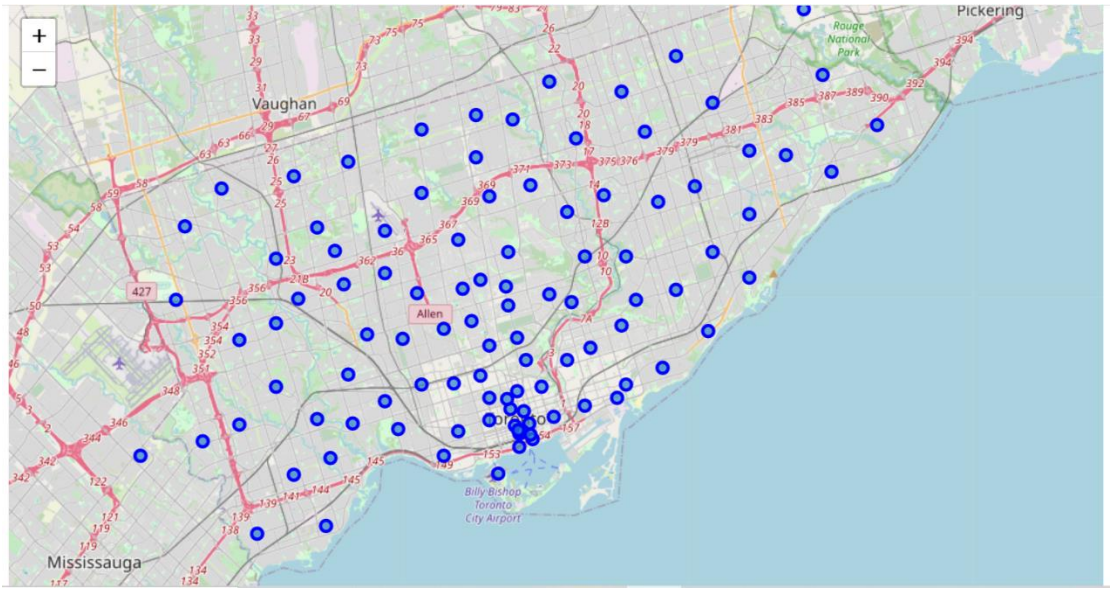

Figure 7

Then we only have to focus on the main part of Toronto. By creating the function of finding the nearby venues, we can get the numbers of nearby venues in a cluster. With the function "return_most_common_venues", we can get the most common venues in each cluster as follows.

	Neighborhood	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	11th Most Common Venue	12th Com V
0	Berczy Park	Coffee Shop	Cocktail Bar	Farmers Market	Bakery	Beer Bar	Restaurant	Seafood Restaurant	Cheese Shop	Eastern European Restaurant	Japanese Restaurant	Irish Pub	Shoppir Mall
1	Brockton, Parkdale Village, Exhibition Place	Café	Coffee Shop	Breakfast Spot	Pet Store	Stadium	Burrito Place	Restaurant	Climbing Gym	Performing Arts Venue	Bakery	Nightclub	Conven Store
2	Business reply mail Processing Centre, South Central Letter Processing Plant Toronto	Light Rail Station	Park	Garden Center	Smoke Shop	Farmers Market	Fast Food Restaurant	Brewery	Burrito Place	Restaurant	Auto Workshop	Skate Park	Pizza Pl
3	CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	Airport Terminal	Airport Lounge	Airport Service	Sculpture Garden	Rental Car Location	Coffee Shop	Boat or Ferry	Harbor / Marina	Bar	Airport Gate	Airport Food Court	Airport
4	Central Bay Street	Coffee Shop	Italian Restaurant	Sandwich Place	Café	Burger Joint	Salad Place	Thai Restaurant	Bubble Tea Shop	Yoga Studio	Indian Restaurant	Seafood Restaurant	Japanes Restaur

First, cluster 3 is not proper obviously because its most common venues are about airport. Cluster 0 is more like a residential area with different types of restaurants, pharmacies and markets. Second, cluster 1 and 2 are very similar with sports places such as gyms, stadiums and yoga studios, and different restaurants and bars. However, cluster 2 is more ideal because there are entertainments like comic shop and skate park in cluster 2. What's more, venues in

cluster 4 are very diversified too with vatious type of restaurants and snack stores. Thus, cluster 2 and cluster 4 are suitable to spend the time on weekend.

3.3 *Check and choose a proper cluster*

The K-clusters method can be used to check the clusters of Toronto.

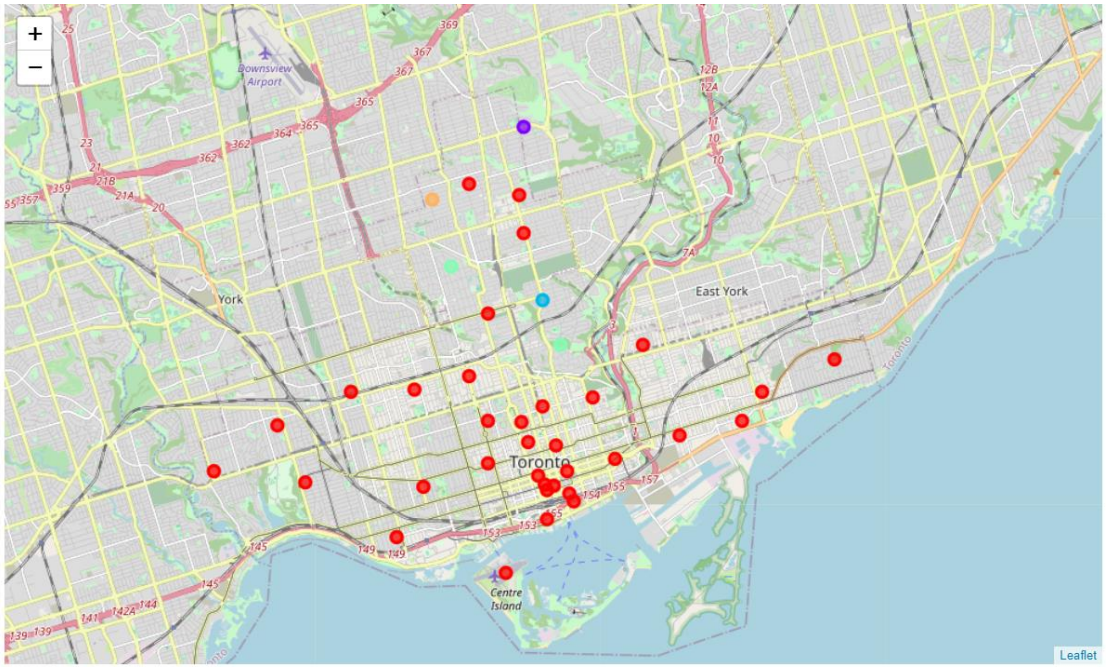

Figure 8

Finally, we can check the distribution of cluster 2 and 4.

	Borough	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	11th Most Common Venue	12th M Comm Ve
8	Central Toronto	2	Lawyer	Tennis Court	Trail	Wine Shop	Doner Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Donut Shop	Department Store	Eastern Europe Restau

	Borough	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	11th Most Common Venue	12th Con V
22	Central Toronto	4	Pool	Garden	Wine Shop	Deli / Bodega	Ethiopian Restaurant	Escape Room	Electronics Store	Eastern European Restaurant	Donut Shop	Doner Restaurant	Dog Run	Distri Cente

As we chose cluster 2 as the place to go, cluster 2 is in Central Toronto as the output shown above. Furthermore, cluster 4 is also a good option with different types of entertainments in Downtown Toronto.

4. Conclusion

In conclusion, if a student in university of Toronto wants to spend the time on weekend, Scaddabush Italian Kitchen & Bar is recommended because of its close distance and good ratings. Then Central Toronto and Downtown Toronto are recommended to visit.