Capstone Project--Solving Problems of University Students

Huimin Jiang

Dec 30, 2020

1. **Description**
   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. 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.

1. **Evaluation of Restaurants**
   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.

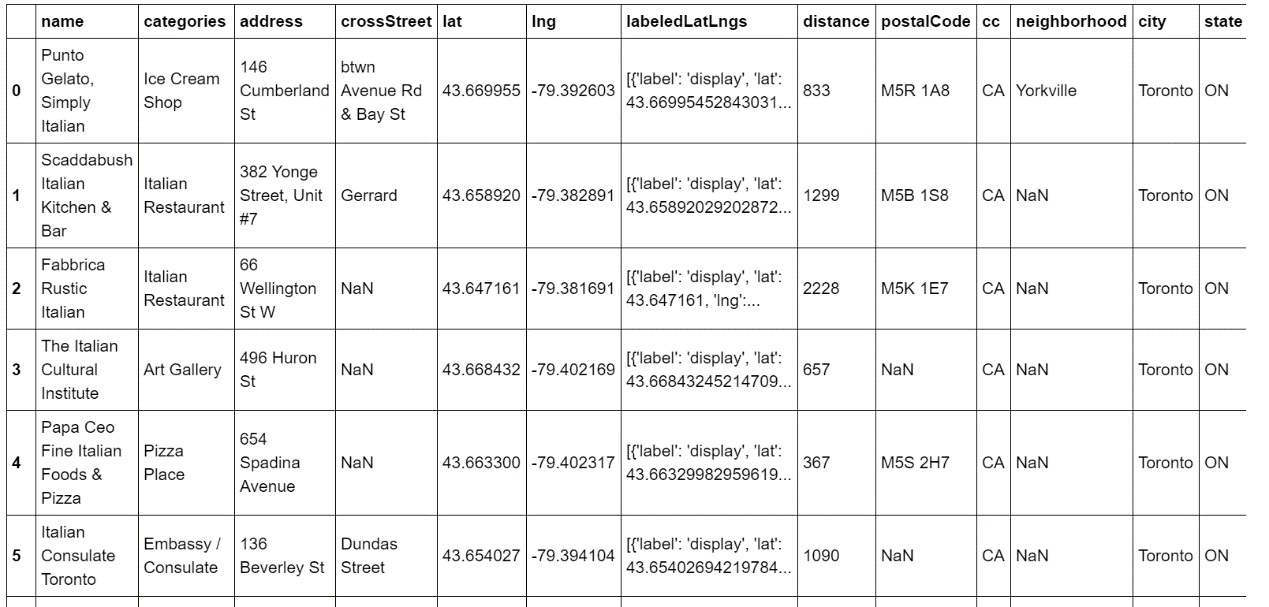


Figure 1

* 1. 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):

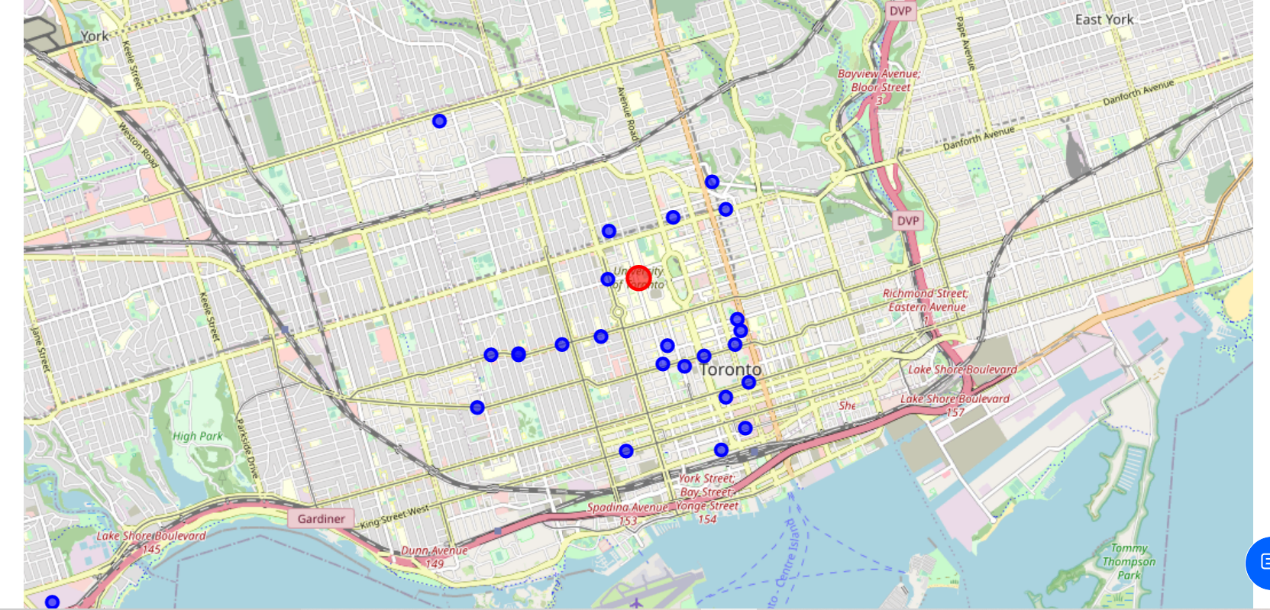


Figure 2

* 1. Further analysis

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

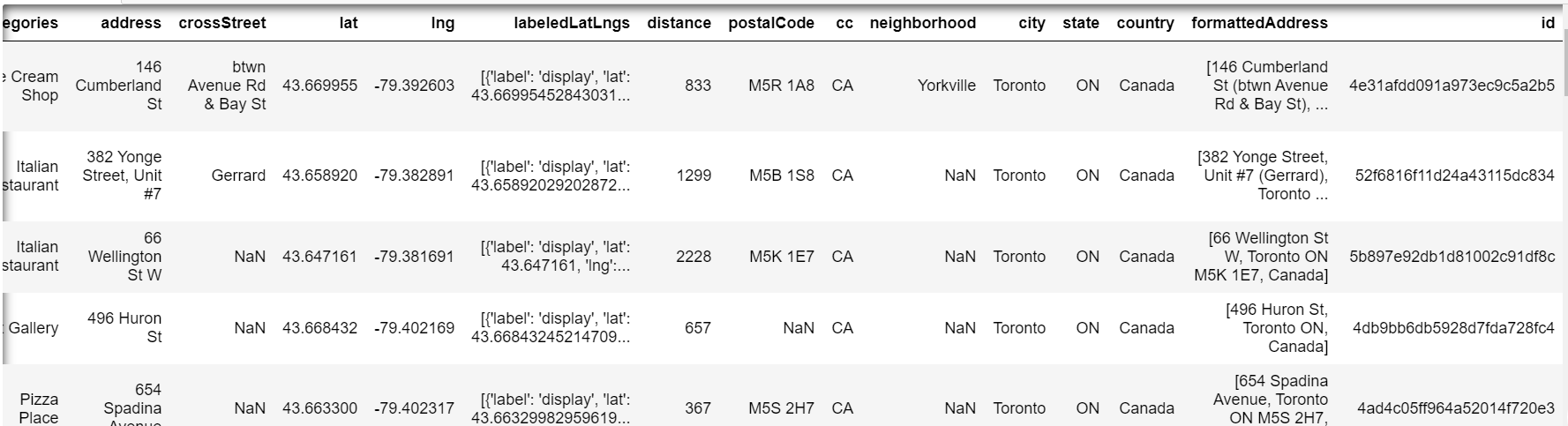


Figure 3

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):

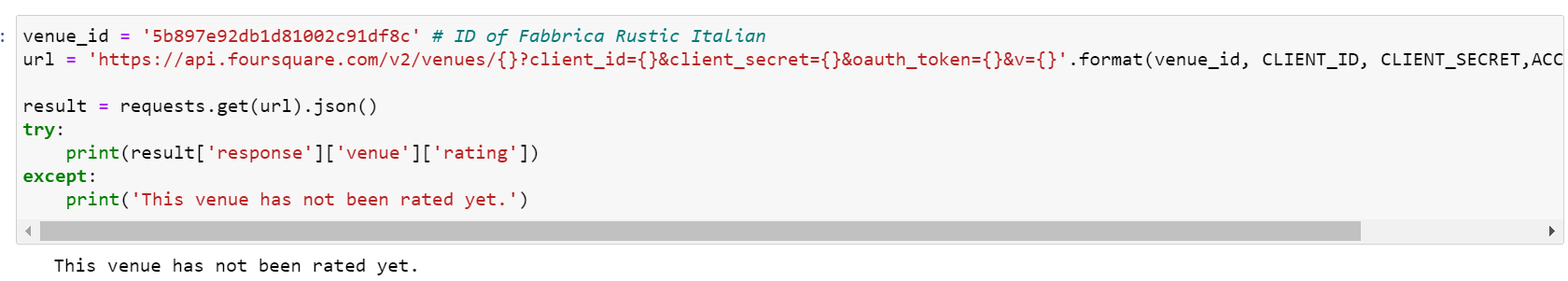


Figure 4

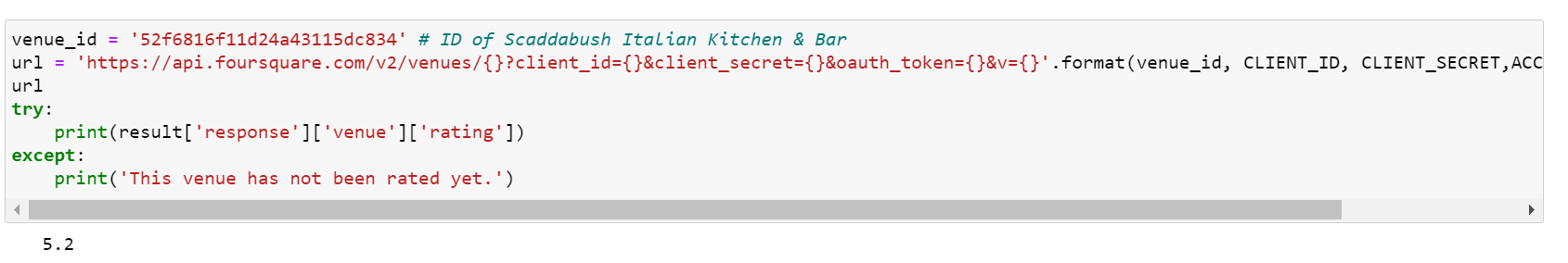


Figure 5

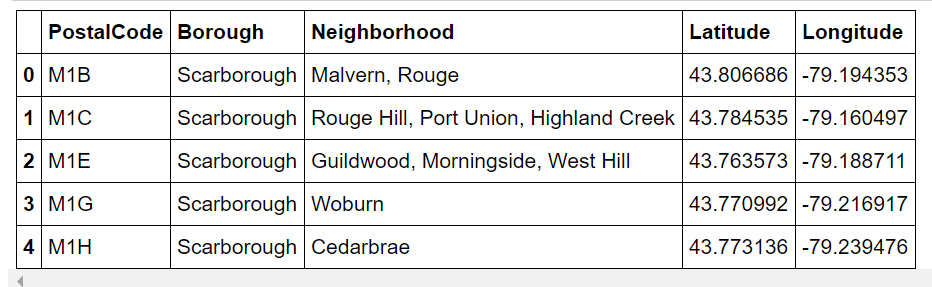


Figure 6

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.

1. **Evaluation of Places in Toronto**
   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.



* 1. Clustering Toronto areas

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

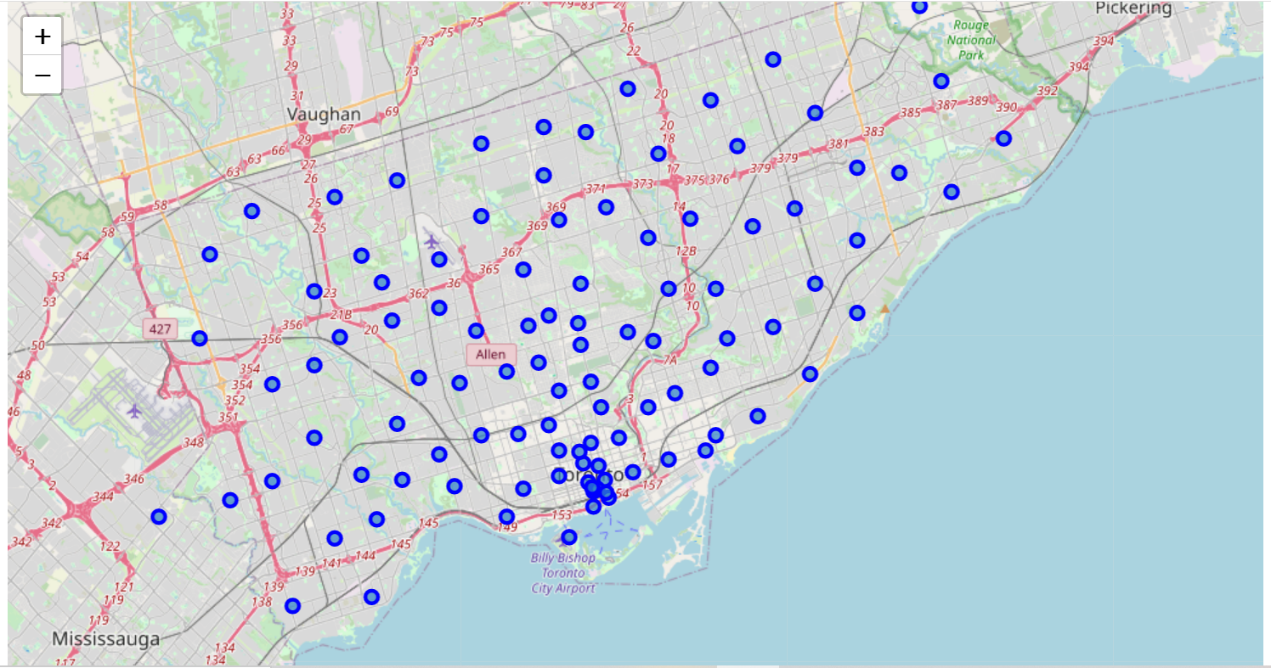


Figure 7

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.



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.

* 1. Check and choose a proper cluster

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

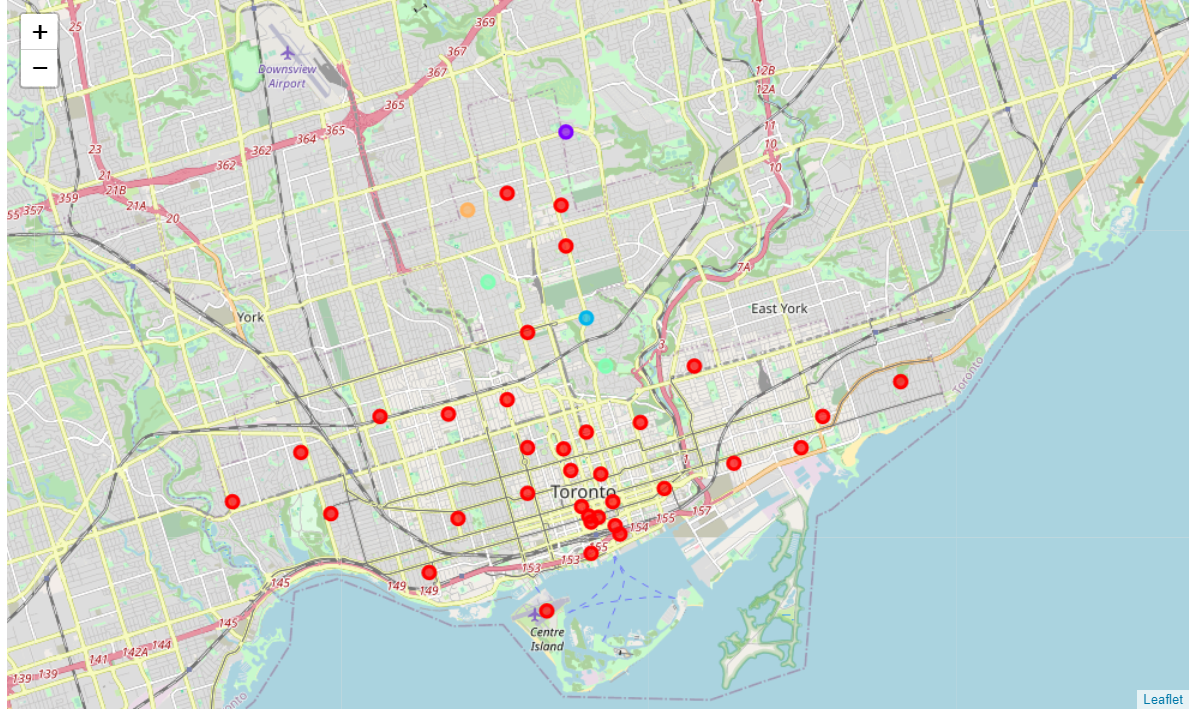


Figure 8

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





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.

1. **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.