Predicting the Best Location for Business Facility

Akshdeep Kaler

August 24, 2020

Introduction/Business Problem

A multi-national company wants to set up the manufacturing facility for Restaurant Equipment and Supply in the neighborhood of Toronto. The company's management wants the perfect location which is closer to restaurant of diverse types. They want diverse customers to get more orders.

To meet their requirements, they gave the data scientist a task to find the neighborhood with diverse restaurant types.

Data Scientist team gathered the data and perform K-means Clustering to give prediction on the diversity of the neighborhood according to the Postal Codes of each neighborhood.

Data Acquisition and Cleaning

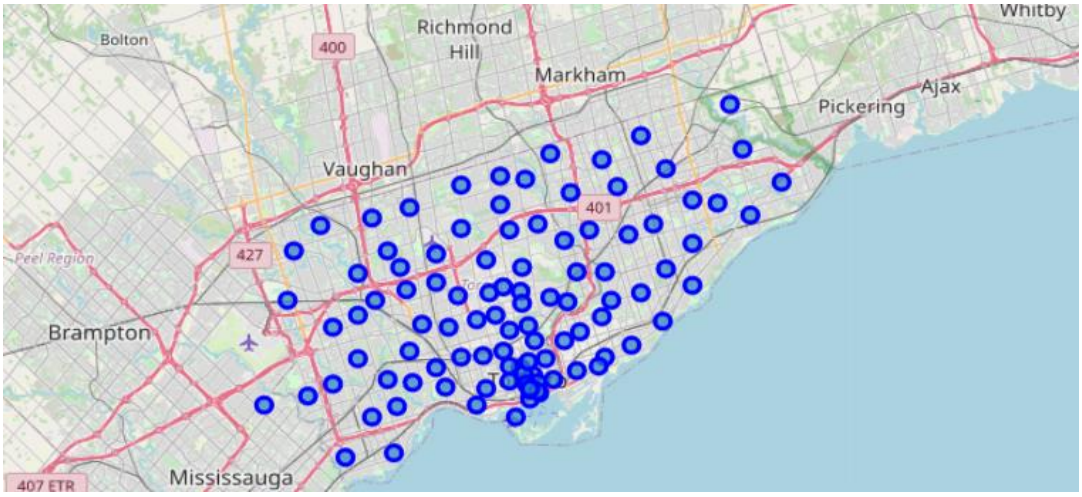
Data downloaded or scraped from Wikipedia were combined into one table. There were missing values in Borough and Neighborhood. Those rows were dropped. Second data which was possessed through Foursquare location search, was used to explore the restaurants in neighborhood of Toronto.

Data feature of 'Postal Code' was used to combine the two table from Wikipedia and Foursquare location to get diverse types of restaurant according to the 'Postal Code' of the neighborhood. Top ten nearest restaurant were listed according to Longitudinal and Latitudinal axis of the location of the Neighborhood.

Exploratory Data Analysis

The gathered data from Wikipedia was the first step towards the search operation of better place to establish the business facility. The data gives out the postal code of the required neighborhood in Toronto city. After cleaning the data I plotted the map to get the better view of the neighborhood.

Fig. 1



The blue clusters give the location of the neighborhoods in Toronto city. (in fig. 1)

As the data from Wikipedia has limited features of the city, it does not gives out the details about the restaurant located around these neighborhood. For that reasons Foursquare API is used to get more information about the neighborhood of Toronto city. Foursquare API gave enormous amount of information. The required information in then cleaned and one hot encoding is applied. After encoding, we got top ten restaurants near to the given location.

Fig.2

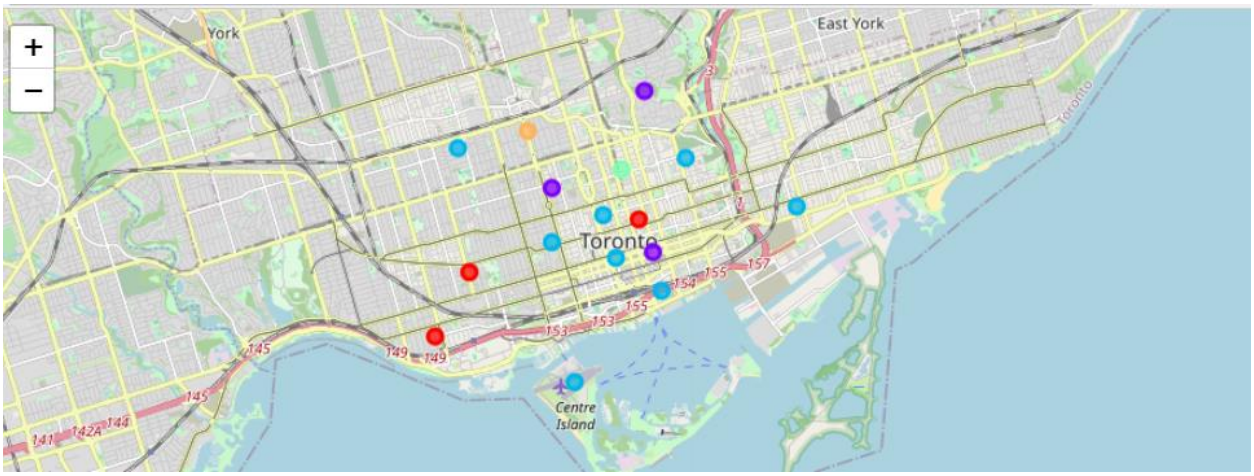
	Postal Code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	M4M	Vietnamese Restaurant	Wine Bar	Bar	Breakfast Spot	Caribbean Restaurant	Chinese Restaurant	Dim Sum Restaurant	Diner	Event Space	Indian Restaurant
1	M4W	Restaurant	Wine Bar	Event Space	Bar	Breakfast Spot	Caribbean Restaurant	Chinese Restaurant	Dim Sum Restaurant	Diner	Indian Restaurant
2	M4X	Chinese Restaurant	Wine Bar	Vietnamese Restaurant	Bar	Breakfast Spot	Caribbean Restaurant	Dim Sum Restaurant	Diner	Event Space	Indian Restaurant
3	M4Y	Thai Restaurant	Wine Bar	Event Space	Bar	Breakfast Spot	Caribbean Restaurant	Chinese Restaurant	Dim Sum Restaurant	Diner	Indian Restaurant
4	M5B	Breakfast Spot	Diner	Wine Bar	Vietnamese Restaurant	Bar	Caribbean Restaurant	Chinese Restaurant	Dim Sum Restaurant	Event Space	Indian Restaurant

Predictive Modelling

In this project, we are helping the customer to establish their business. according to them they want diverse restaurant types around their facility. So that they can touch more types of customers in their business of interest which is supplying and servicing Restaurant appliances and equipment.

For prediction the better location I used K-Means Clustering. We use K-Means clustering to divide the Postal Codes and Neighborhood with similar characteristics into clusters. In this case, we divided them into five clusters. This method gave the variety of restaurants in each cluster. Each cluster has its own characteristics. In this example, characteristics are represented the type of restaurants. We analyze through each cluster to attain best possible location for new facility according to the requirements of the customer.

Fig. 3



Conclusions

Every cluster presents the specific characteristics of its own. According to requirement of diverse restaurants type, Cluster 3 is very much suitable. Cluster 3 has many Borough and to select the best one I selected the postal code with substantial number of neighborhoods. In this case, M5V is best option for setting up the new facility.

```
Postal Code                                                          M5V
Borough                                                 Downtown Toronto
Neighbourhood            CN Tower, King and Spadina, Railway Lands, Har...
Latitude                                                         43.6289
Longitude                                                       -79.3944
Cluster Labels                                                         2
1st Most Common Venue                                          Wine Bar
2nd Most Common Venue                                        Restaurant
3rd Most Common Venue                                               Bar
4th Most Common Venue                                  Indian Restaurant
5th Most Common Venue                                    Thai Restaurant
6th Most Common Venue                                       Noodle House
7th Most Common Venue                                         Nightclub
8th Most Common Venue                          New American Restaurant
9th Most Common Venue                                 Korean Restaurant
10th Most Common Venue                              Vietnamese Restaurant
Name: 12, dtype: object
```