The Battle of Neighborhoods
(IBM DATA SCIENCE CAPSTONE PROJECT Applied Data Science Course)

Opening a Chinese Restaurant in the city of AMSTERDAM

Analyst: Joey Tse

25/05/2021

Content

1. Introduction

This report is an IBM Data Science Professional Certification – Capstone Project named The Battle of Neighborhoods. Recall that I have explored New York City and the city of Toronto and segmented and clustered their neighborhoods. Here, I would introduce another multicultural city – Amsterdam, Netherlands, to do neighborhoods map analysis. (Amsterdam would be compared with New York, US and Toronto, Canada, to see their similarity and dissimilarity among these financial capitals.)

Amsterdam is a multicultural city with diverse cultural backgrounds' people. It is capital of Netherlands with a population of 872,680 within the city proper, 1,558,755 in the urban area and 2,480,394 in the metropolitan area. [1] This largest city in the Netherlands contents approximately 178 different cultural backgrounds, hosting different cultural events throughout the year that shows its wide-world view.

1.1 Business Problem

Our client is an investor who is interested in investing a restaurant in Amsterdam. He have approached us to study the market and determine a neighborhood in order to starting his business there. The client, a constracter, also want a suitable place to setup his office to run the business. So, our main goal is to suggest a right place to him, which satisfies both conditions by analyzing data using various data science techniques.

The conditions of the neighborhood would be:

- Location of Chinese restaurant should be in tourism area for attracting tourists or families to have a meal.
- Location of the office should be near the restaurant, which give administration support and store the raw food.

1.2 Target Audience

For this project, we would like to do these three things.

- Some overseas Chinese would like to visit to the restaurant since the taste of hometown is nostalgic.
- Apart from overseas Chinese, we also target on the foreigners because Amsterdam is multicultural. So, our business may be a combination of eastern and western culture.
- Since we may consider the location without other similar types of Chinese food shop, we may concentrate on the tourists who want an amazing try and the local families for weekdays gathering time.

[1] https://en.wikipedia.org/wiki/Amsterdam

2. Data

2.1 Source of data and Usage

First, our analyst team would do web scraping from a Wikipedia page contents a list of Neighbourhoods of Amsterdam. (https://en.wikipedia.org/wiki/Category:Neighbourhoods_of_Amsterdam). We extract the dataset by using requests library to handle requests and BeautifulSoup package to extract the source.

Second, we would like to use Foursquare API to get data of venue for those neighborhoods. Foursquare is a database of more than 105 million places worldwide and an API that enables location data for over 150,000 high technology brands or GPS developers. We import geopy.geocoders module to convert an address into latitude and longitude values of Amsterdam for further map plotting and getting all of the neighborhoods.

Third, we would use sklearn.cluster package import K-means clustering to identify which cluster (area/ location) is more suitable to start a Chinese restaurant. K-means clustering is an unsupervised learning, aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. We then visualize the results though mapping.

2.2 Dataset

1	Amsteldorp
2	Amsterdam Oud-West
3	Amsterdam Oud-Zuid
4	Amsterdam Science Park
5	Apollobuurt
6	Betondorp
7	Bijlmermeer
8	Binnenstad (Amsterdam)
9	Bos en Lommer
10	Buiksloot
11	Buikslotermeer
12	Buitenveldert
13	Bullewijk
14	Burgwallen Nieuwe Zijde
15	Burgwallen Oude Zijde
16	Chass?buurt
17	Cruquiuseiland
18	Czaar Peterbuurt
19	Dapperbuurt
20	De Aker

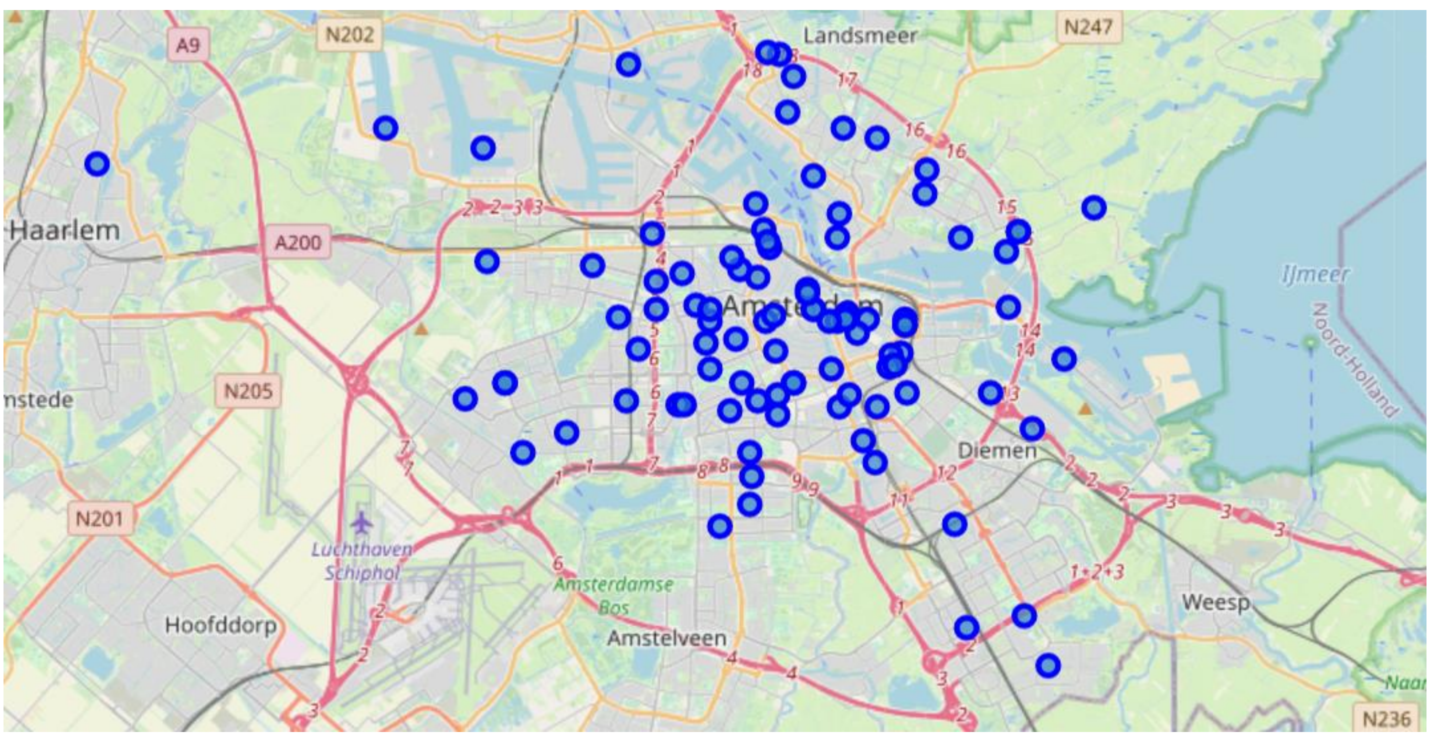
* head of 20 column, total 105 neighborhoods

3 Methodology

First of all, we explored and loaded the dataset in dataframe format. We did web scraping using requests library and BeautifulSoup package, it is an one column of list of neighborhoods in the city of Amsterdam. We would like to get their geographical coordinates as we had to plot a map and get an overall view on what the graph looks like. Then, we used Foursquare API to get data of venue in forms of latitude and longitude.

	Neighborhood	Latitude	Longitude
0	Admiralenbuurt	52.372752	4.856359
1	Amsteldorp	52.339680	4.918740
2	Amsterdam Oud-West	52.365390	4.870220
3	Amsterdam Oud-Zuid	52.352350	4.877880
4	Amsterdam Science Park	52.354300	4.958010
5	Apollobuurt	52.350294	4.867990
6	Betondorp	52.423405	4.833395
7	Bijlmermeer	52.307031	4.969744
8	Binnenstad (Amsterdam)	52.369930	4.907880
9	Bos en Lommer	52.379190	4.851740

After setting up FourSquare credentials, we got the latitude and longitude of Amsterdam and plotted geographic map with makers.



Then, we set a limit number of venues (100) with radius 500 returned by FourSquare API to create URL, this could be processed to run a get_ catrgory_type function for those venues. There were 45 venues returned by Foursquare. We explored neighborhoods in Amsterdam to get the name, latitude, longitude and category of venues for all neighborhoods.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Admiralenbuurt	52.372752	4.856359	Sapporo Ramen Sora	52.371294	4.855144	Ramen Restaurant
1	Admiralenbuurt	52.372752	4.856359	Deli-caat	52.371221	4.856056	Deli / Bodega
2	Admiralenbuurt	52.372752	4.856359	Rein Cityspa	52.371217	4.855969	Spa
3	Admiralenbuurt	52.372752	4.856359	Café Cook	52.371208	4.852792	Pub
4	Admiralenbuurt	52.372752	4.856359	Maz Mez	52.371231	4.857968	Lebanese Restaurant

Next, for each of the neighborhoods, we did feature extraction by one-hot encoding. This method is to make each feature to be a category that belongs to corresponding venue, and convert to binary number, 1 means this category is found in the venue and 0 is vis verse. We used grouping to show the frequency of each category of restaurants in each city district

	Neighborhood	Zoo Exhibit	Accessories Store	Afghan Restaurant	African Restaurant	American Restaurant	Aquarium	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entert
0	Admiralenbuurt	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
1	Amsteldorp	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
2	Amsterdam Oud-West	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.011364	0.0000
3	Amsterdam Oud-Zuid	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.016949	0.000000	0.000000	0.016
4	Amsterdam Science Park	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
5	Apollobuurt	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.040000	0.000000	0.000000	0.0000
6	Betondorp	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.111111	0.000000	0.000000	0.0000
7	Bijlmermeer	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
8	Binnenstad (Amsterdam)	0.000000	0.00	0.000000	0.016949	0.000000	0.000000	0.016949	0.000000	0.016949	0.000000	0.000000	0.0000
9	Bos en Lommer	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
10	Buiksloot	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000

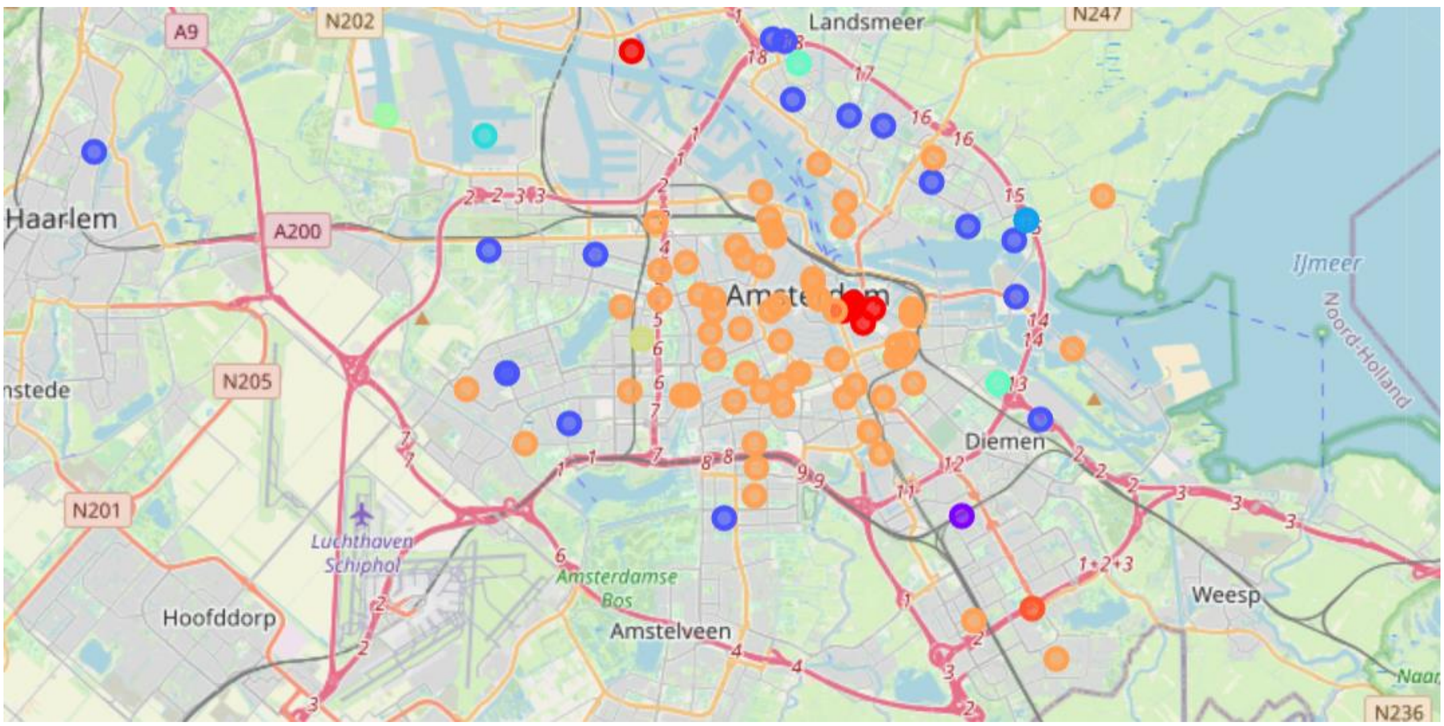
After that, I could finally run an unsupervised machine learning algorithm, more specifically, a k-means clustering algorithm from the scikit-learn package. One could use the elbow method to systematically define the k value, I simply chose k up to 10.

4 Results

In the below table, we have been assigned ten different cluster labels from 0 to 9. The nine columns of common venue showed different types of food shops and restaurants.

	Neighborhood	Latitude	Longitude	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
0	Admiralenbuurt	52.372752	4.856359	8	Restaurant	Café	Deli / Bodega	Snack Place	Bar	Supermarket	Coffee Shop	Pub	Falafel Restaurant
1	Amsteldorp	52.339680	4.918740	8	Hotel	Italian Restaurant	Brasserie	Sports Club	Dutch Restaurant	French Restaurant	Café	Furniture / Home Store	Gastropub
2	Amsterdam Oud-West	52.365390	4.870220	8	Café	Coffee Shop	Italian Restaurant	Restaurant	Bar	Hotel	Grocery Store	Gym / Fitness Center	Yoga Studio
3	Amsterdam Oud-Zuid	52.352350	4.877880	8	Restaurant	Hotel	Bar	Bakery	Plaza	Cosmetics Shop	French Restaurant	Juice Bar	Concert Hall
4	Amsterdam Science Park	52.354300	4.958010	5	Bus Stop	Coffee Shop	Spa	Convenience Store	Restaurant	Ethiopian Restaurant	Doner Restaurant	Drugstore	Dutch Restaurant

We used the cluster labels to show the city districts marked with a cluster-specific color on a map, again using folium:



There are quite a number of bubbles for all neighborhoods. We may focus on the bubbles colored in light-green, purple and light-blue.

These are clusters showing lower numbers of food shops and restaurants in those districts. It might have greater potential to run a Chinese restaurant and setup its office nearby because it has less competitiveness with others. The locations of these clusters are mainly concentrated on the northern and eastern area of capital Amsterdam. They are connected by railway, which is convenience to people to travel around. We are no need to worry about the customer stream.

Here are the restaurant type concentrations for the city of Amsterdam:

```
# cluster 1
a_merged.loc[a_merged['Cluster Labels'] == 0, a_merged.columns[[0] + [3] + list(range(4, a_merged.shape[1]))]]
```

	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
6	Betondorp	0	Furniture / Home Store	Market	Museum	Historic Site	Dutch Restaurant	Art Gallery	Seafood Restaurant	Empanada Restaurant	Doner Restaurant	Drugstore
8	Binnenstad (Amsterdam)	0	Bar	Restaurant	Hostel	Breakfast Spot	Café	Pizza Place	Museum	Italian Restaurant	Hotel	History Museum
17	Cruquiuseiland	0	Bar	Restaurant	Hostel	Breakfast Spot	Café	Pizza Place	Museum	Italian Restaurant	Hotel	History Museum
27	Floradorp	0	Bar	Restaurant	Hostel	Breakfast Spot	Café	Pizza Place	Museum	Italian Restaurant	Hotel	History Museum
32	Grachtengordel	0	Bar	Restaurant	Hostel	Breakfast Spot	Café	Pizza Place	Museum	Italian Restaurant	Hotel	History Museum
41	Kadijken	0	Bar	Restaurant	Hostel	Breakfast Spot	Café	Pizza Place	Museum	Italian Restaurant	Hotel	History Museum
44	KNSM Island	0	Bar	Restaurant	Hostel	Breakfast Spot	Café	Pizza Place	Museum	Italian Restaurant	Hotel	History Museum

```
# cluster 2
a_merged.loc[a_merged['Cluster Labels'] == 1, a_merged.columns[[0] + [3] + list(range(4, a_merged.shape[1]))]]
```

	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
94	Venserpolder	1	Restaurant	Zoo	Event Service	Dog Run	Doner Restaurant	Drugstore	Dutch Restaurant	Electronics Store	Empanada Restaurant	Ethiopian Restaurant

```
# cluster 3
a_merged.loc[a_merged['Cluster Labels'] == 2, a_merged.columns[[0] + [3] + list(range(4, a_merged.shape[1]))]]
```

	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 Com Veni
10	Buiksloot	2	Supermarket	Bus Stop	Park	Restaurant	Drugstore	Shopping Mall	Turkish Restaurant	Café	Bakery	Emp Rest
11	Buikslotermeer	2	Supermarket	Convenience Store	Clothing Store	Electronics Store	Hotel	Seafood Restaurant	Sporting Goods Shop	Snack Place	Shopping Mall	Bus
26	Eendracht (Amsterdam)	2	Soccer Field	Bus Stop	Baseball Field	Cafeteria	General Entertainment	Stadium	Snack Place	Event Space	Garden	Athl Spo
31	Gouden Reael	2	Gym / Fitness Center	Supermarket	Playground	Shop & Service	French Restaurant	Liquor Store	Tennis Court	Dog Run	Bakery	Drug
37	Indische Buurt	2	Snack Place	Playground	Video Store	Bath House	Shoe Store	Bus Stop	Chinese Restaurant	Grocery Store	Indonesian Restaurant	Ice C Sho

```python
# cluster 4
a_merged.loc[a_merged['Cluster Labels'] == 3, a_merged.columns[[0] + [3] + list(range(4, a_merged.shape[1]))]]
```

	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
56	Nieuwendam	3	Garden	Dog Run	Plaza	Skating Rink	Zoo	Empanada Restaurant	Doner Restaurant	Drugstore	Dutch Restaurant	Electronics Store
91	Tuindorp Nieuwendam	3	Garden	Dog Run	Plaza	Skating Rink	Zoo	Empanada Restaurant	Doner Restaurant	Drugstore	Dutch Restaurant	Electronics Store

```python
# cluster 5
a_merged.loc[a_merged['Cluster Labels'] == 4, a_merged.columns[[0] + [3] + list(range(4, a_merged.shape[1]))]]
```

	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
98	Westelijk Havengebied	4	Boat or Ferry	Zoo	Event Space	Drugstore	Dutch Restaurant	Electronics Store	Empanada Restaurant	Ethiopian Restaurant	Event Service	Exhibit

```python
# cluster 6
a_merged.loc[a_merged['Cluster Labels'] == 5, a_merged.columns[[0] + [3] + list(range(4, a_merged.shape[1]))]]
```

	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
4	Amsterdam Science Park	5	Bus Stop	Coffee Shop	Spa	Convenience Store	Restaurant	Ethiopian Restaurant	Doner Restaurant	Drugstore	Dutch Restaurant	Electronics Store
50	Molenwijk (Amsterdam)	5	Bus Stop	Supermarket	Bookstore	Furniture / Home Store	Shopping Mall	Pharmacy	Zoo	Empanada Restaurant	Drugstore	Dutch Restaurant

```python
# cluster 7
a_merged.loc[a_merged['Cluster Labels'] == 6, a_merged.columns[[0] + [3] + list(range(4, a_merged.shape[1]))]]
```

	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
79	Ruigoord	6	Coffee Shop	Bar	Zoo	Event Service	Doner Restaurant	Drugstore	Dutch Restaurant	Electronics Store	Empanada Restaurant	Ethiopian Restaurant

```python
# cluster 8
a_merged.loc[a_merged['Cluster Labels'] == 7, a_merged.columns[[0] + [3] + list(range(4, a_merged.shape[1]))]]
```

	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
72	Overtoomse Veld	7	Supermarket	Fast Food Restaurant	Gym	Drugstore	Sandwich Place	Nightclub	Dog Run	Doner Restaurant	Dutch Restaurant	Electronics Store

5 Conclusion

5.1 Recommendations

It is recommended Nieuwendam, Tuindorp Nieuwendam and Westelijk Havengebied (in Cluster 4 and 5), maybe Tuindorp Oostzaan and Zeeburgereiland are also avaliable to the client. Those are some traffic stations and entertainment places for next decision. He may select the most preferrable place to open his Chinese restaurant and setup his business office nearby.

5.2 Discussion

For the K-means clustering, the cluster number 10 is quite a lot to handle in clustering analysis. Neat time, I may try to control the cluster number into 5 with six clusters in order to merge similar neighborhoods together.

For the category identification, I may not consider both different kinds of food shops, such as café shop, bar, plaza place, etc., and restaurants. I would like to focus on all types of restaurants only. This may better select the place with least restaurants.

End of Report