

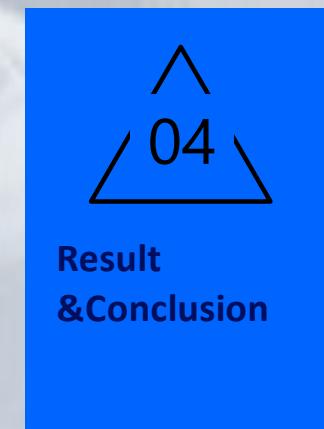
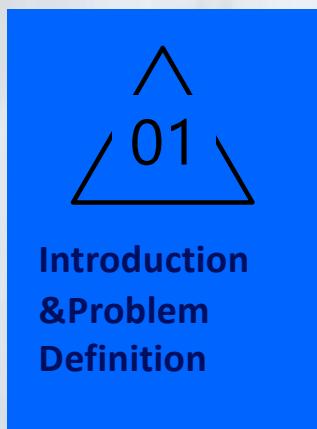


IBM Applied Data Science Capstone Project Final Report

**Finding Best Place to date with friends in
Beijing, China**

*Hong
Feb. 2020*

Contents



Introduction & Problem Definition

Introduction & Problem Definition: In a multicultural city and huge city like Beijing, China. it might be daunting to find out which places are best to date your friends comfortably and relax after working time , considering the traffic issues of the venues, maybe near the subway station is better in a high traffic city, Not only that, one needs to look into different factors like shopping easy , restaurants, cafes good eating taste and so on.

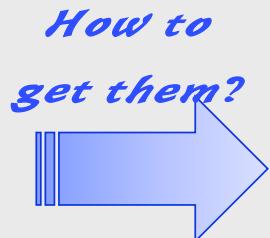
Here can take me to a subway station that I often use as location information and I go to foursquare to explore the types of venues in the street, intuitively sort the data from the places I used to go, extract features, and help me to visualize the characteristics of the data

For The Audiences: *Who lives in Beijing and have subway transportation hobby want to find a suitable venue to have a gathering with his friends based on difference purpose/interesting*

Data Collection & Preparation

To solve the problem, we need to find the following data:

- List of Neighborhoods/ subway station and their properties
- Latitude and Longitude of the neighborhoods/ subways.
- Venue data of the neighborhoods



- First** find the list of neighborhoods from csv file located on the IBM Cloud Object Storage '<https://s3.eu-geo.objectstorage.service.networklayer.com>' and get coordinates of each neighborhoods including the average of all latitudes and longitudes. This would be needed in order to plot the neighborhood clusters.
- Second** find the venue data, Foursquare API was used. It would show the most popular venues in each neighborhoods while using the central coordinates of the respective neighborhoods.
- Third** this data is all marked and unstructure data , and we should do is to standardize and visualize for overview understanding. Maybe we should pay more attention to Feature selection and the processing of missing values.

Methodology & Tools

The following tools used:

- 01 The Foursquare API was used to obtain venue data.
- 02 The Folium package was used to plot Neighborhoods on the map
- 03 The KMeans module from sklearn package was used to cluster the data
- 04 The json package was used to open and read the geojson file.
- 05 IBM Cloud Object Storage - Python SDK from ibm_boto3 to get the coordinates of Neighborhoods
- 06 The .read_csv from pandas library to scrape the data from a .csv file

Methodology & Tools

Analytic Approach

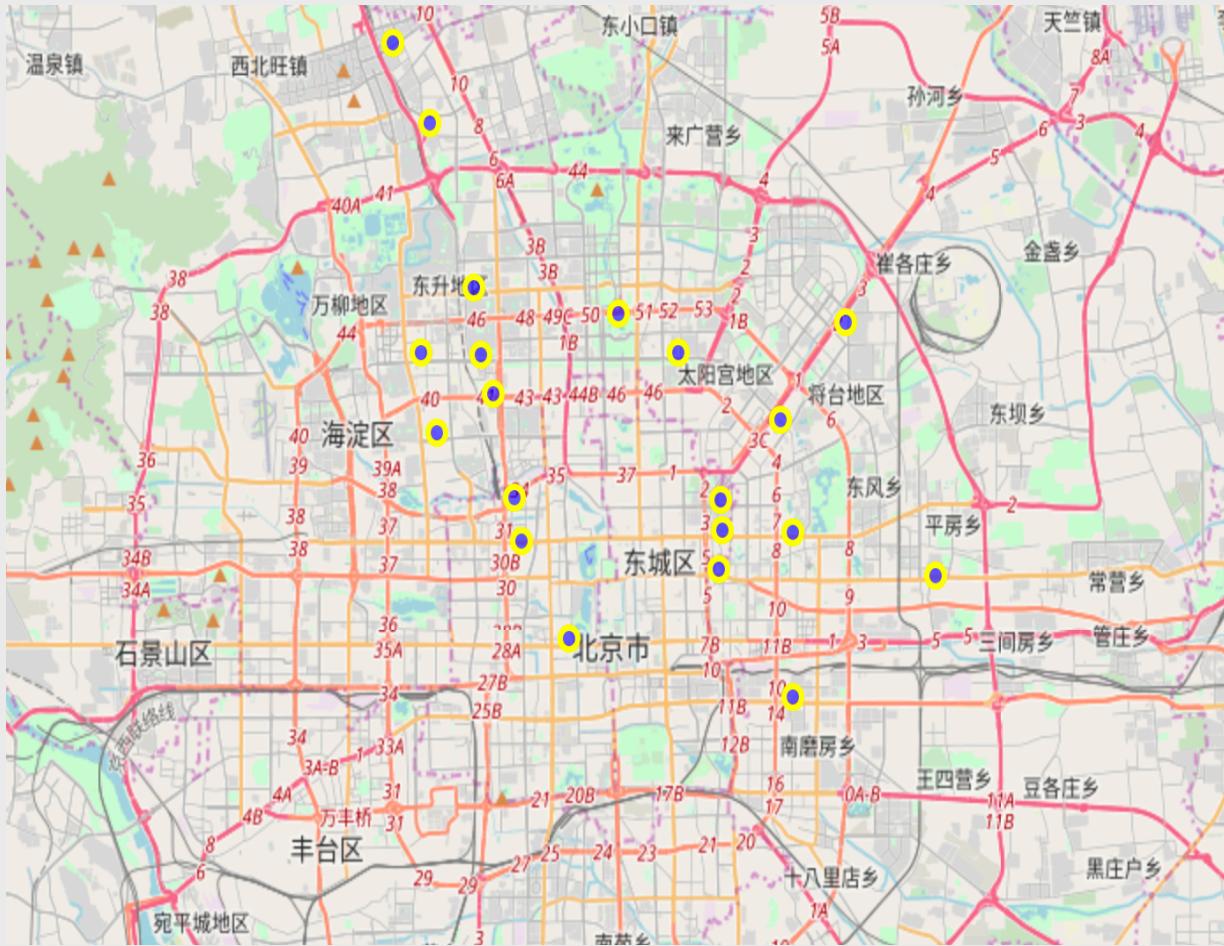
01

First of all, the data we get is the comprehensive data based on the latitude and longitude from foursquare. This data contains many dimensions. The problem to be solved is to explore the feature in cluster. There are not limited standardization for weather it is right or false. Because we do not need to predict at this stage. We just want to find out the result of clusters from the data in entertainment facilities. The algorithm selected in this project is K-means

02

Clustering is the division of data into groups such that data points in the same group are more similar than data points in other groups. In short, clustering is the division of data points with similar characteristics into groups, that is, clusters. The goal of the K-means algorithm is to find a group in the data, the number of groups being represented by the variable K. Each data point is assigned to one of the K groups by an iterative operation based on the characteristics provided by the data.

Result & Conclusion



Subway Stations Overview

I have selected about 20 subway stations where people visit very often and explore the possibilities around the subways. .

Result & Conclusion

Venues							
(1769, 7)							
[6]:	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	XIERQI	40.052243	116.306144	李记潮汕砂锅粥	40.041352	116.335103	Cantonese Restaurant
1	XIERQI	40.052243	116.306144	Shantou Baheli Hai's Beef Restaurant (汕头八合里海记牛肉店)	40.040046	116.333920	Cantonese Restaurant
2	XIERQI	40.052243	116.306144	Starbucks (星巴克)	40.052671	116.296030	Coffee Shop
3	XIERQI	40.052243	116.306144	Starbucks (星巴克)	40.046375	116.292217	Coffee Shop
4	XIERQI	40.052243	116.306144	Haidilao Hot Pot (海底捞火锅)	40.027357	116.305631	Hotpot Restaurant
...
1764	WEIGONGCUN	39.957765	116.323136	Sculpting in Time (雕刻时光)	39.955710	116.305167	Café
1765	WEIGONGCUN	39.957765	116.323136	麻辣诱惑 Spice Spirit	39.977089	116.309981	Chinese Restaurant
1766	WEIGONGCUN	39.957765	116.323136	Carrefour (家乐福)	39.979649	116.307309	Supermarket
1767	WEIGONGCUN	39.957765	116.323136	雕刻时光 Sculpting In Time	39.978723	116.308185	Café
1768	WEIGONGCUN	39.957765	116.323136	Costa Coffee (咖世家)	39.979337	116.306323	Coffee Shop

1769 rows × 7 columns

```
[8]: print('There are {} uniques categories.'.format(len(Venues['Venue Category'].unique())))
```

There are 141 uniques categories.

Categories Overview

Using Foursquare API to obtain the facilities venue data around each subway stations/ Neighborhoods , there are :

- 1769 venues item results.
- 141 unique categories from all the returned venues

Result & Conclusion

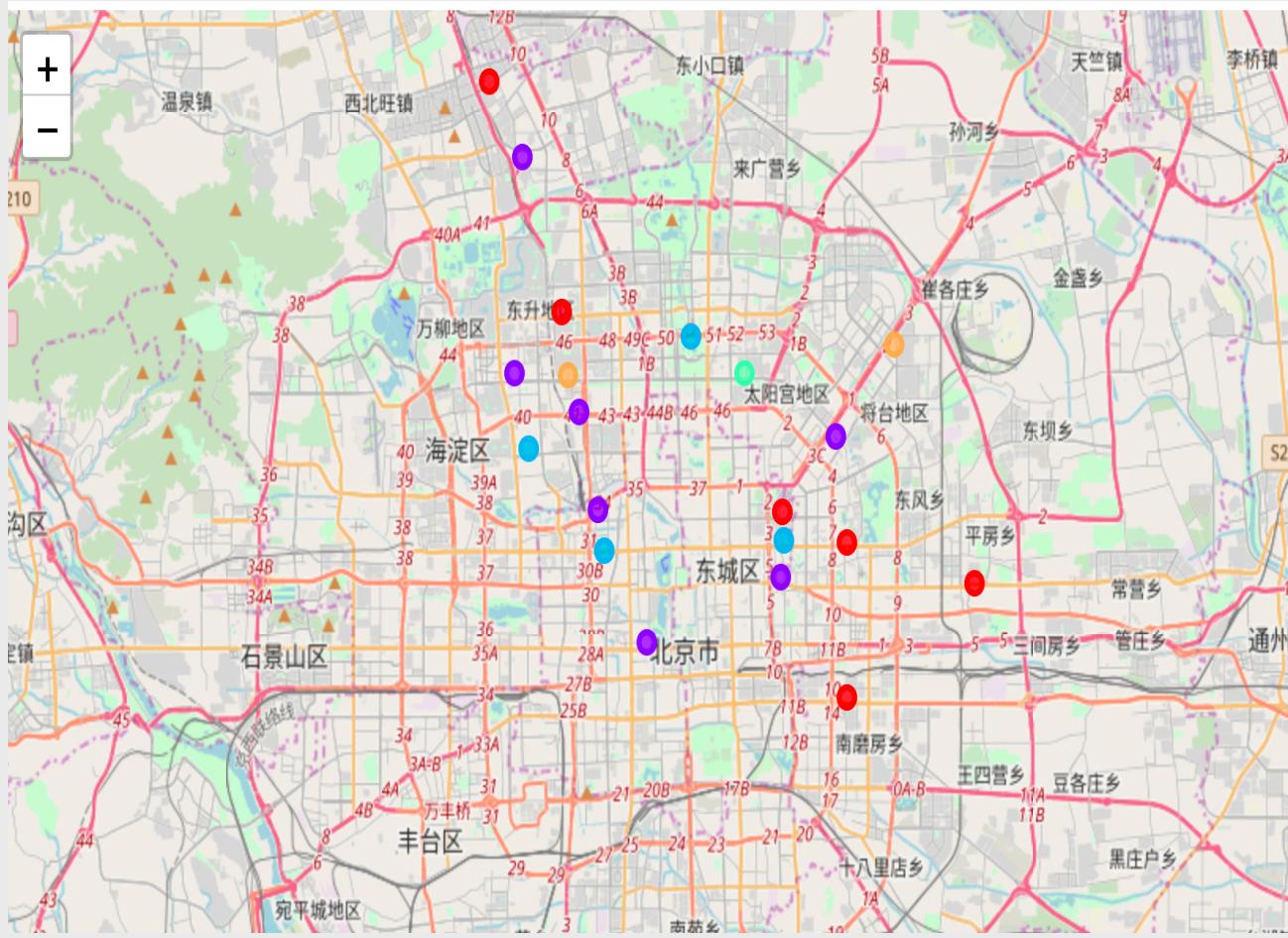
13]:

	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
0	AOTIZHONGXIN	Chinese Restaurant	Hotel	Coffee Shop	Fast Food Restaurant	Pizza Place	Hotpot Restaurant	Café	Park	Szechuan Restaurant	New American Restaurant
1	CHAoyANGMEN	Hotel	Shopping Mall	Chinese Restaurant	Café	Dumpling Restaurant	Italian Restaurant	Bar	Brewery	Peking Duck Restaurant	French Restaurant
2	CHEGONGZHUANG	Chinese Restaurant	Coffee Shop	Fast Food Restaurant	Hotel	Pizza Place	Café	Historic Site	Hotpot Restaurant	Department Store	Szechuan Restaurant
3	DAZHONGSI	Fast Food Restaurant	Pizza Place	Coffee Shop	Café	Chinese Restaurant	Sandwich Place	Bar	Clothing Store	Hotel	Hotpot Restaurant
4	DONGSISHITIAO	Hotel	Chinese Restaurant	Brewery	Japanese Restaurant	Café	Shopping Mall	Bar	Pizza Place	Szechuan Restaurant	Dumpling Restaurant
5	DONGZHIMEN	Hotel	Chinese Restaurant	Japanese Restaurant	Café	Brewery	Coffee Shop	Shopping Mall	Pizza Place	Szechuan Restaurant	Dumpling Restaurant
6	HAI DIAN HUANGZHUANG	Chinese Restaurant	Fast Food Restaurant	Café	Sandwich Place	Coffee Shop	Pizza Place	Bakery	Bar	Korean Restaurant	Xinjiang Restaurant
7	HUI XIN XI JIENANKOU	Fast Food Restaurant	Chinese Restaurant	Coffee Shop	Hotel	Pizza Place	Hotpot Restaurant	Asian Restaurant	Multiplex	Shopping Mall	Park
8	QINGNIANLU	Supermarket	Coffee Shop	Hotel	Clothing Store	American Restaurant	Noodle House	Park	Farm	Electronics Store	Department Store
9	SANYUANQIAO	Hotel	Japanese Restaurant	Bakery	Italian Restaurant	Park	Chinese Restaurant	Café	Shopping Mall	Cocktail Bar	Brewery

Top 10 Features

- Changed the data of string for categories into number and group rows by neighborhood and by taking the mean of the frequency of occurrence of each category
- Selected TOP 10 features for Neighborhood clustering

Result & Conclusion



Result

- Used TOP 10 features to Neighborhood cluster in 5 clusters using KMeans module.
- Visualized the clustering result in the Map
- Here we didn't cover the best K using the Elbow Analysis

Result & Conclusion

Checked the details of 5 clusters. I found : 2 clusters (cluster 4&cluster 5) only contain 1 or 2 items , So we just discuss the rest 3 clusters (cluster1,cluster2 & cluster 3) and get below result:



01

*Asia style
cluster*

In the first cluster most of them are:
Coffee Shop, Café,
Chinese Restaurant, Fast Food Restaurant,
Dumpling Restaurant,
Japanese Restaurant,
Asian Restaurant

02

*Diversity
cluster*

In the Second cluster most of them are:
Hotel, Chinese Restaurant, Japanese Restaurant, Italian Restaurant, French Restaurant, Pizza Place, Xinjiang Restaurant, Shopping Mall, Korean Restaurant, Fast Food Restaurant, Cocktail Bar, Fast Food Restaurant, Peking Duck Restaurant, Brewery, Asian Restaurant

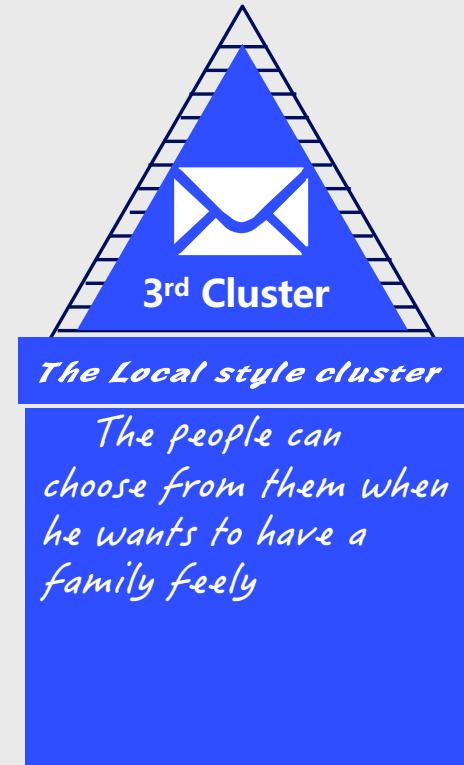
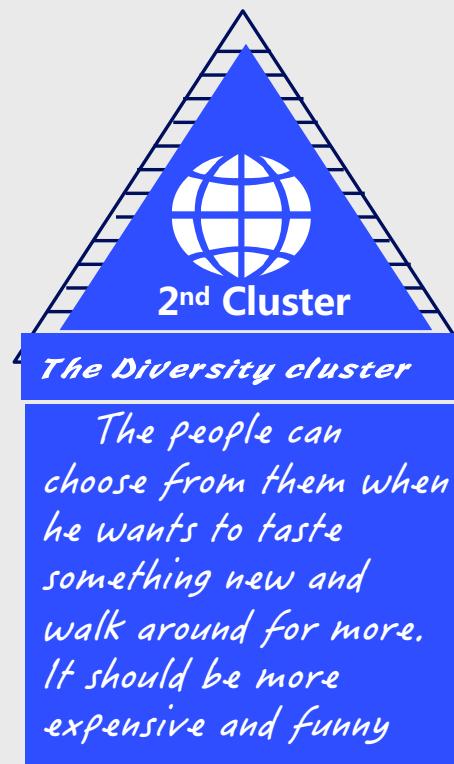
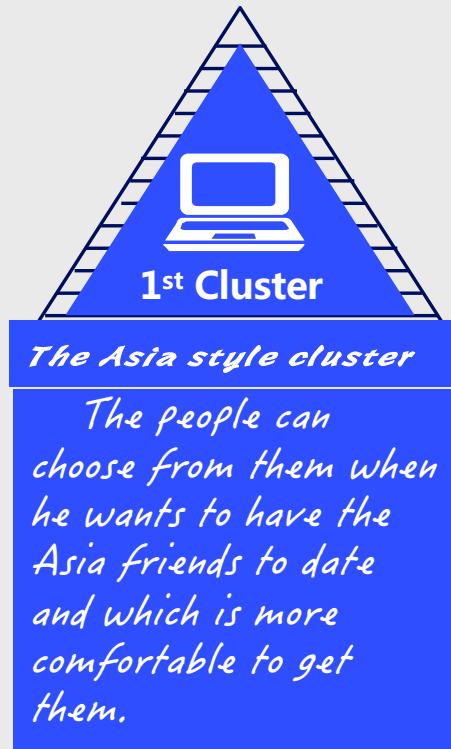
03

*local style
cluster*

In the Third cluster most of them are:
Chinese Restaurant, Hotpot Restaurant, Coffee Shop, Fast Food Restaurant, Szechuan Restaurant, Café

Result & Conclusion

The principle of making a choice



*Thanks and
Welcome to China!*

