

Applied Data Science Capstone Final Report

1. Introduction

As a foreigner working/studying in Taipei (Taiwan) with little knowledge of Chinese, I face a lot of difficulties in understand the city and its people. With the power of Python, I will try to explore how things locate in this city.

I will make two attempts: one for all venue category and one for restaurant only (because we all love food).

Target audiences: anyone who likes to explore Taipei, especially from the viewpoint of a foreigner with little language to communicate.

2. Data

2.1. Neighborhood locations

The easiest location of each areas is the post office representing that neighborhood. In this part I use pandas to read the data from Postal location of Taipei city from <https://www.post.gov.tw/post/internet/index.jsp> .

The result of the data without geocoding is taipei_add.csv. Due to some difficulties to find API geocode, I do not present the code here to get the latitude and longitude of the venues. Also, the length of the data is cut down to 48 from more than 150. The latitude and longitude data are shown in taipei_add_geocode.csv

2.2. Foursquare data

After getting the latitude and longitude of various neighborhood in Taipei, will use Foursquare data to get the most popular venues in each neighborhood.

3. Method

We explore the Taipei city data by following these steps:

- + Get the location data
- + Geocode it into latitude and longitude
- + Use Foursquare to examine each neighborhood
- + Cluster the neighborhood using KNN

In this work, I will use this method for two attempts:

- + Apply for all locations with all venue category.
- + Apply for all locations with categories contain the string "Restaurant".

4. Result

4.1. Getting the data locations, geocode and test Foursquare API

The data are scraped and geocoded, which are ready for further analysis (Fig. 1).

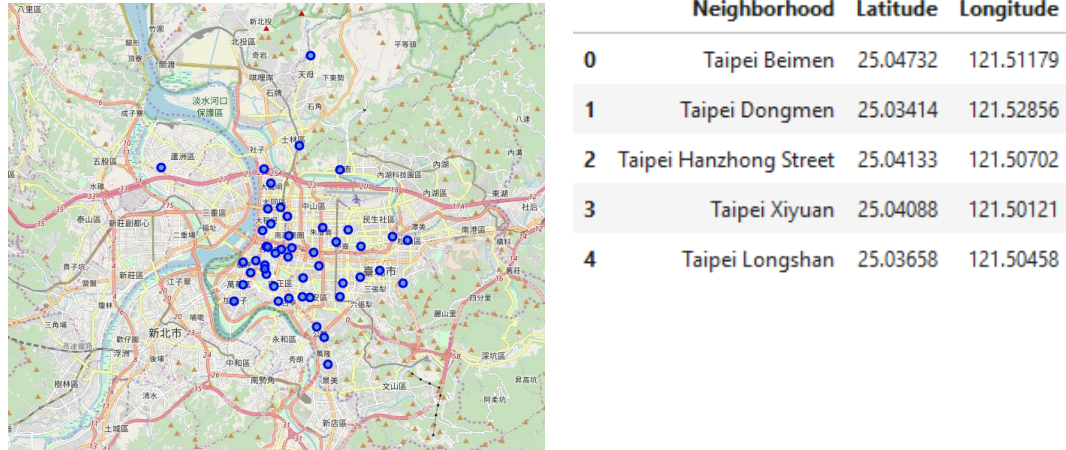


Figure 1. Preview of the neighborhood uses in this project

We test the Foursquare API with one venue and it returns a very promising result (Fig. 2).

```
dict_keys(['meta', 'response'])  
61 venues were returned by Foursquare.
```

	name	categories	lat	lng
0	鄉記豬腳飯	Asian Restaurant	25.046989	121.511049
1	North Gate (台北府城北門)	Historic Site	25.047584	121.511179
2	Heritage Bakery & Cafe	Café	25.045171	121.511824
3	張家清真黃牛肉麵館 Chang's Halal beef Noodles	Noodle House	25.045718	121.510720
4	修圖素食 Vegetarian / Vegan Restaurant	Vegetarian / Vegan Restaurant	25.046702	121.514228

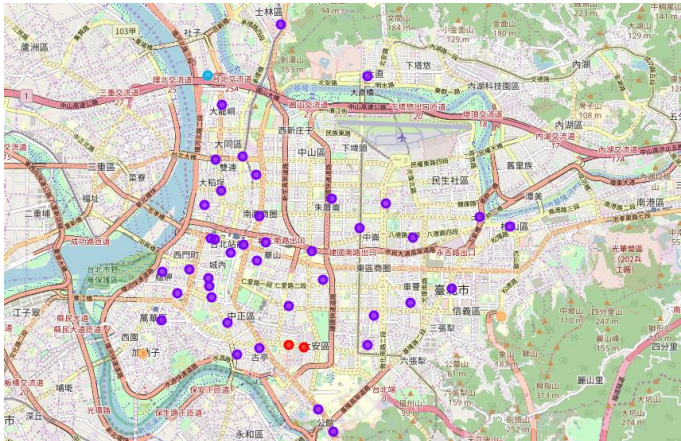
Figure 2. Sample test of the Foursquare API on a single neighborhood.

4.2. KNN clustering using all the data

We used all the data to do the classification. All of the locations have been used as the input for the Foursquare API and then transformed into features ready for KNN classification. The result of the clustering is shown below:

	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Group 1	Café	Noodle House	Sushi Restaurant	Chinese Restaurant	Tea Room
Group 2	Café	Convenience Store	Noodle House	Coffee Shop	Chinese Restaurant
Group 3	Intersection	Burger Joint	Bus Station	Chinese Restaurant	Yunnan Restaurant
Group 4	Fast Food Restaurant	Department Store	Burger Joint	Planetarium	Flea Market
Group 5	Convenience Store	Park	Coffee Shop	Bus Station	Fast Food Restaurant

Figure 3. KNN clustering of all locations

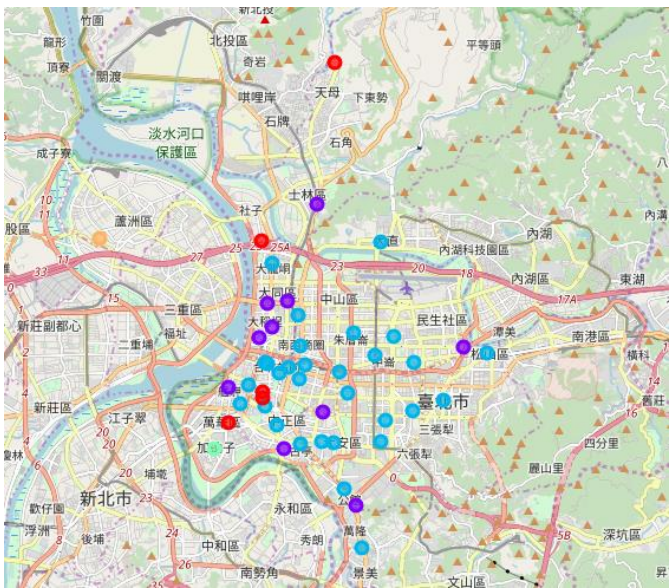


4.3. Clustering the locations based on the restaurant only

We made the second attempt with a filtering of data. In this experiment, we only used the data with venue categories containing the string “Restaurant”. The result is shown below:

	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Group 1	Chinese Restaurant	Japanese Restaurant	Greek Restaurant	Japanese Restaurant	Japanese Curry Restaurant
Group 2	Taiwanese Restaurant	Japanese Restaurant	Hotpot Restaurant	Japanese Restaurant	Greek Restaurant
Group 3	Chinese Restaurant	Chinese Restaurant	Hotpot Restaurant	Asian Restaurant	Fast Food Restaurant
Group 4	Hotpot Restaurant	Yunnan Restaurant	Greek Restaurant	Japanese Restaurant	Japanese Curry Restaurant
Group 5	Fast Food Restaurant	Yunnan Restaurant	Greek Restaurant	Japanese Restaurant	Japanese Curry Restaurant

Figure 4. KNN clustering of restaurant only.



5. Discussion and conclusion

We can see from the first attempt of clustering that Taipei city is filled with Café and restaurant everywhere and they are the most common venues in all places. Only the different clustering groups are located in outer side of the city where other public venues including parks, intersection, etc. are most common.

In the second attempt with only restaurant venue, we have a more interesting clustering of data. Group 1,2 are mostly inner cities areas filled with Chinese Restaurant or Taiwanese restaurant while group 3,4,5 contain more exotic restaurant like Hotpot, Yunnan or Greek Restaurant. This information can be used for travelers to choose their favorite food.

In conclusion, this project has been used to explore the neighborhood of Taipei city. The most distinguish feature is the fertile availability of restaurants in this area. For a classification of restaurant, we can see the clustering in different areas with different types of foods.