Applied Data Science Capstone Final Report

1. Introduction

As a foreigner living and working in Taipei City of Taiwan, with a little knowledge of Chinese language, sometimes I face difficulties in understanding the city, places, locations and its people. With the power of Python, I will try to explore how places are located in the city.

I will make to attempts: one for all venue categories and one for restaurants only.

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

2. Data

The easiest way to find locations of each areas in the post office representing that neighborhood. In this part I use pandas to read the data from postal location of Taipei from official website of Taiwan post.

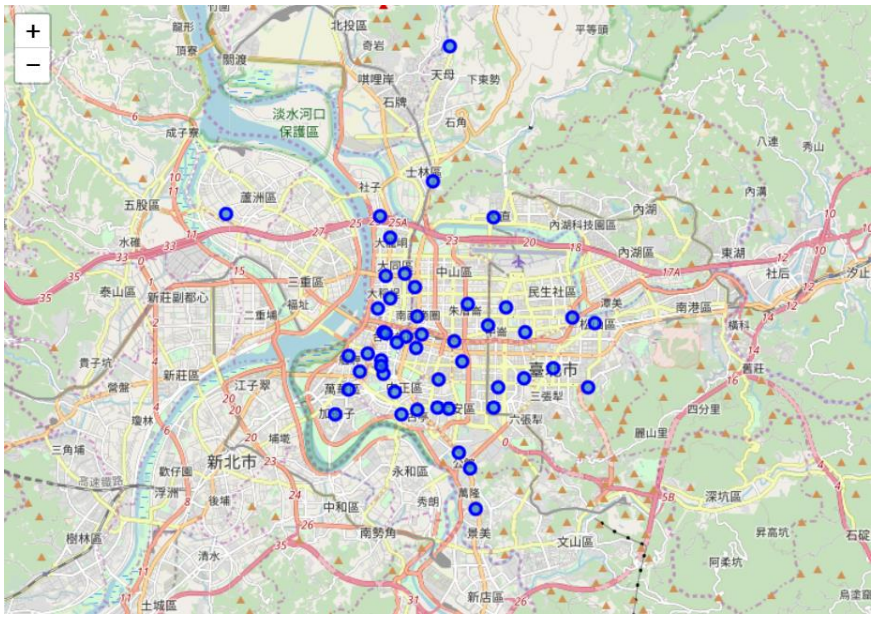
After getting coordinates of various neighborhood in Taipei, I will use Foursquare data to get the most popular venues in each neighborhood.

3. Method

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

4. Result

The data are scraped and geocoded, which are ready for further analysis.



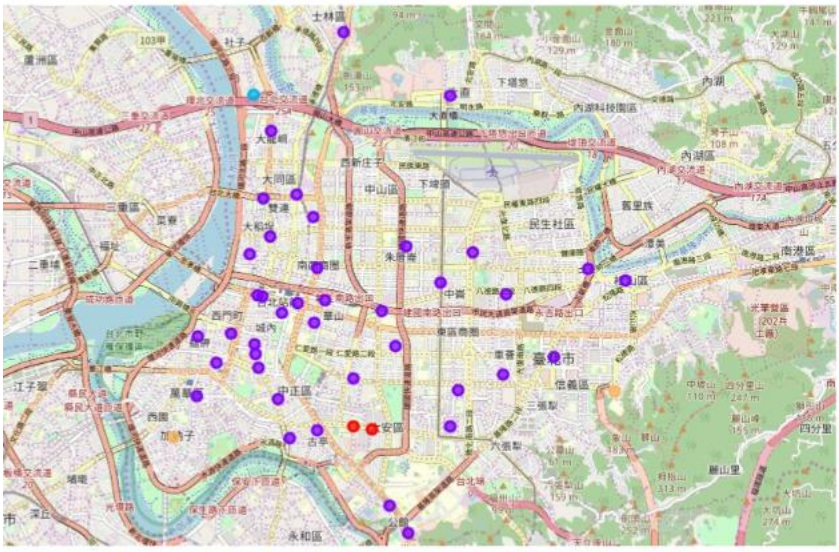
After testing the Foursquare API with one venue, and it returned this result:

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

	name	categories	lat	lng
0	鄭記豬腳飯	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	修圓素食	Vegetarian / Vegan Restaurant	25.046702	121.514228

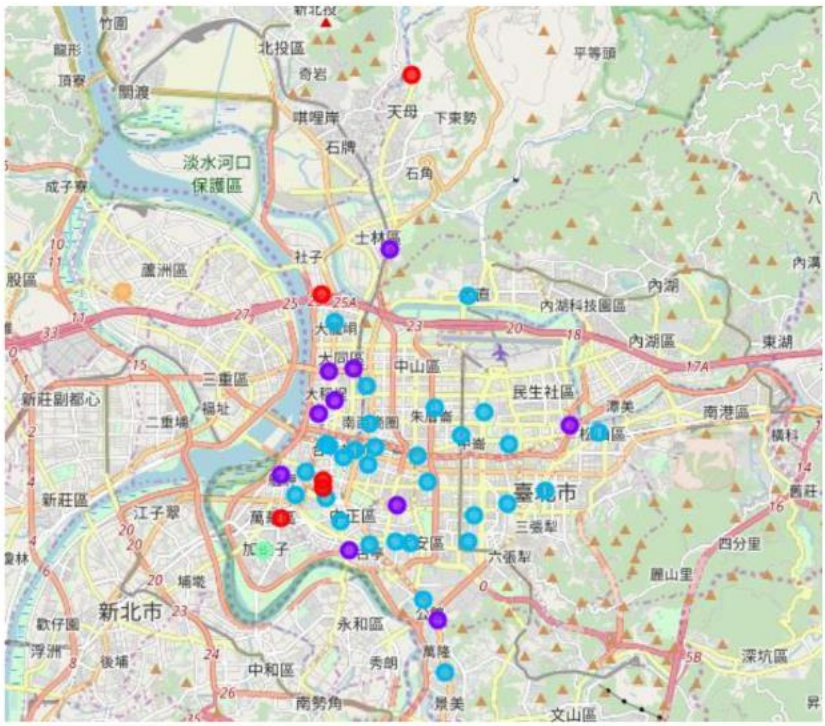
All the data was used to do the classification. All of the locations have been used as the input for the Foursquare API and then transformed into features ready for KNN classification. The result oof the clustering is shown below:

	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



Second attempt was done using a filtering of data. In this case, we only used the data with venue categories containing the string “Restaurant”. The result is 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



5. Conclusion

Can see from the first attempt of clustering that Taipei City is filled with café and restaurants and they are the most common venues in all places. The only difference that the clustering groups are located in outer side of the city where other public venues are more common.

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

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