**Capstone Project - The Battle of Neighborhoods - Taichung version**

This project is a part of IBM data science, you will find in this post an overview of my final capstone project.

In this assignment, I will go through the problem description (**Introduction/Business Problem**), data set preparation and final to analysis and overview these data step by step. Detailed code are given in Github and link can be found at the end of the post.

1. **Introduction/Business Problem**
   1. **Background**

**Taiwan is a famous nation on the world. It was vary small (36,197** km2 **) but with more than 23,600,000 population. Taiwan with the high metropolitan GDP in the world and it belonged a developed country.**

**There was COVID-19 spread in the world but it seems not impact to Taiwan. Here with the best epidemic prevention concept and medical technology so the infection rate and mortality rate is the lowest in the world.**

**With estimated 11,840,000 overseas visitors every year. It’s famous with delicious foods and beautiful landscapes. Many travelers needed to know some leisure and entertainment places when they first visit to Taiwan.**

**That will also attract many immigrations come to Taiwan.**

**Taichung is a second big city in Taiwan, it was smaller than Taipei only. Taichung retained the advantage of Taipei like: convenient transportation/** **convenient life/good job opportunity/medical system…etc. Many immigrations will choose settle down in Taichung.**

* 1. **Business Problem**

**Taichung with 29 different districts, some districts are commercial and some are convenient life. These districts are so difference!!!**

**In this article will help let immigrations to understand and overview the Taichung city. Let them to choose the district which they want to live.**

1. **Data Preparation**
   1. Data Clean and Preprocess:
      * Define CSV Process Class



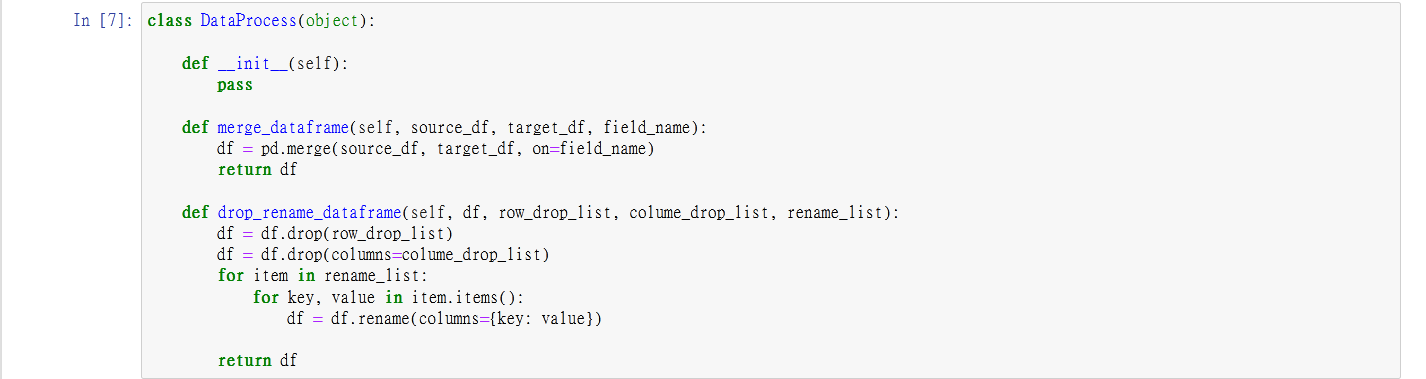
* + - Define Json Process Class



* + - Define Get Data from Web Class



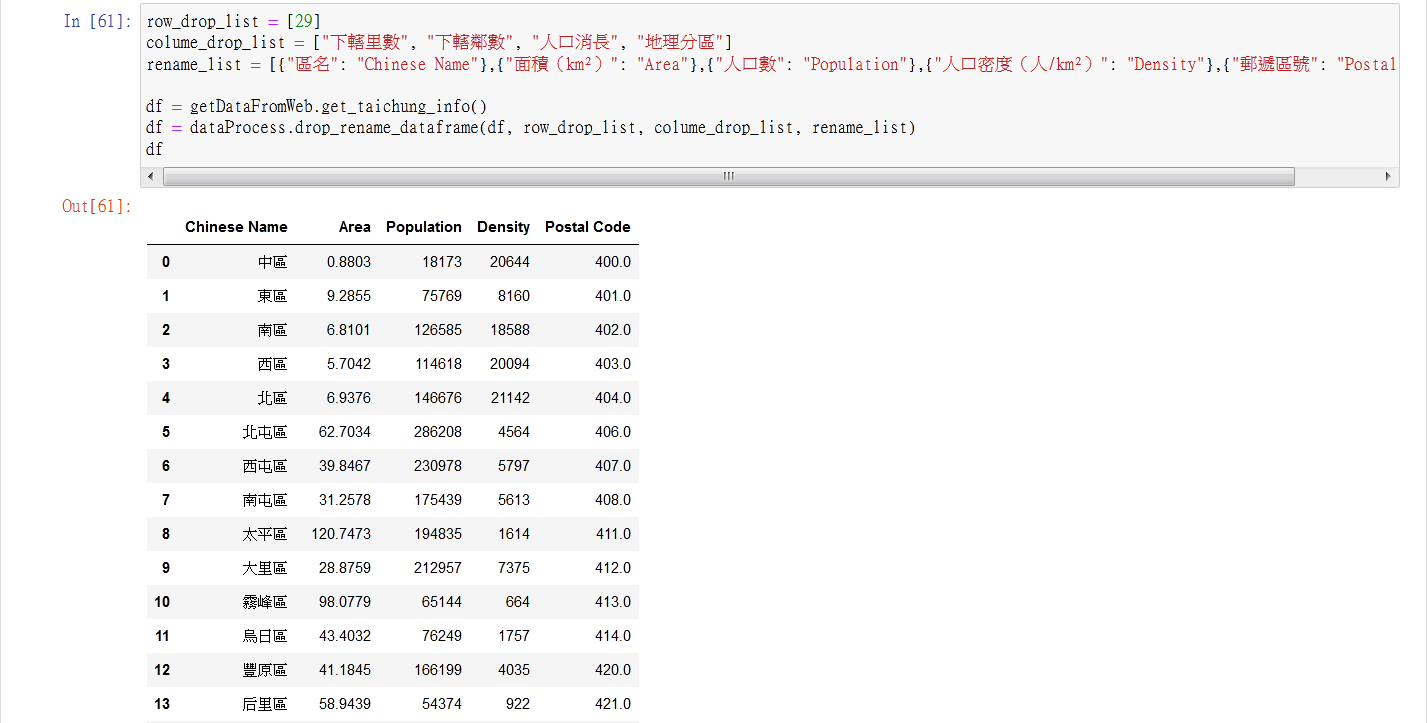
* + - Define DataFrame Process Class

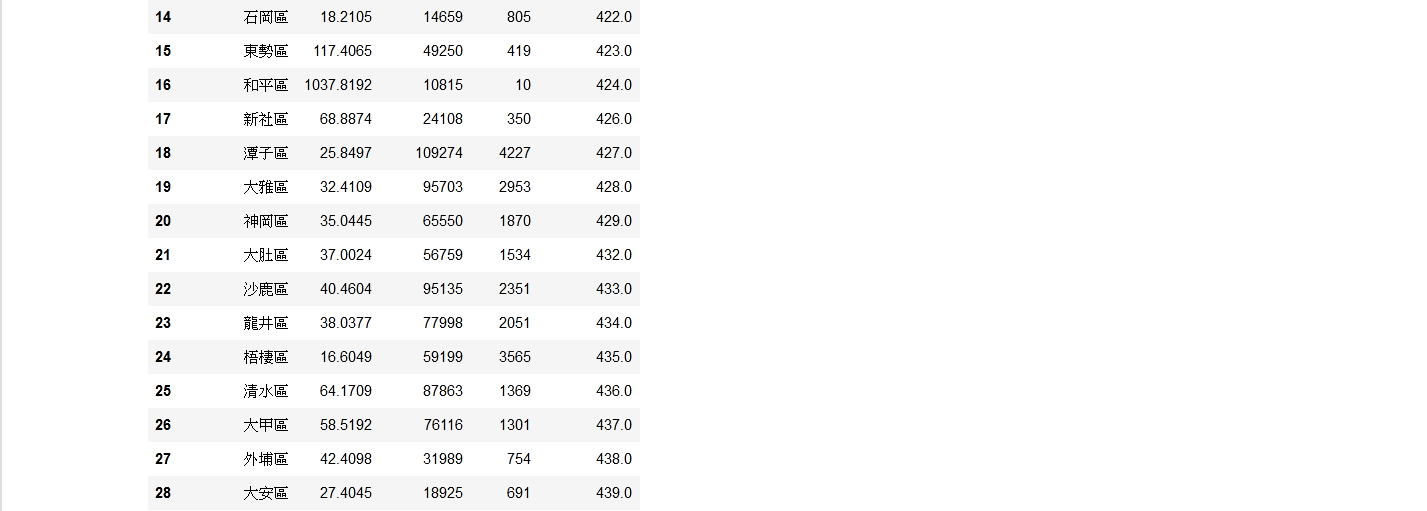


* 1. Taichung wiki:

Data source: <https://zh.wikipedia.org/wiki/臺中市#人口>

Description: This data included 29 **districts in Taichung city. There are some** field we needed like: Chinese Name/Area/Density/Postal Code.





* 1. Latitude and longitude in Taiwan:

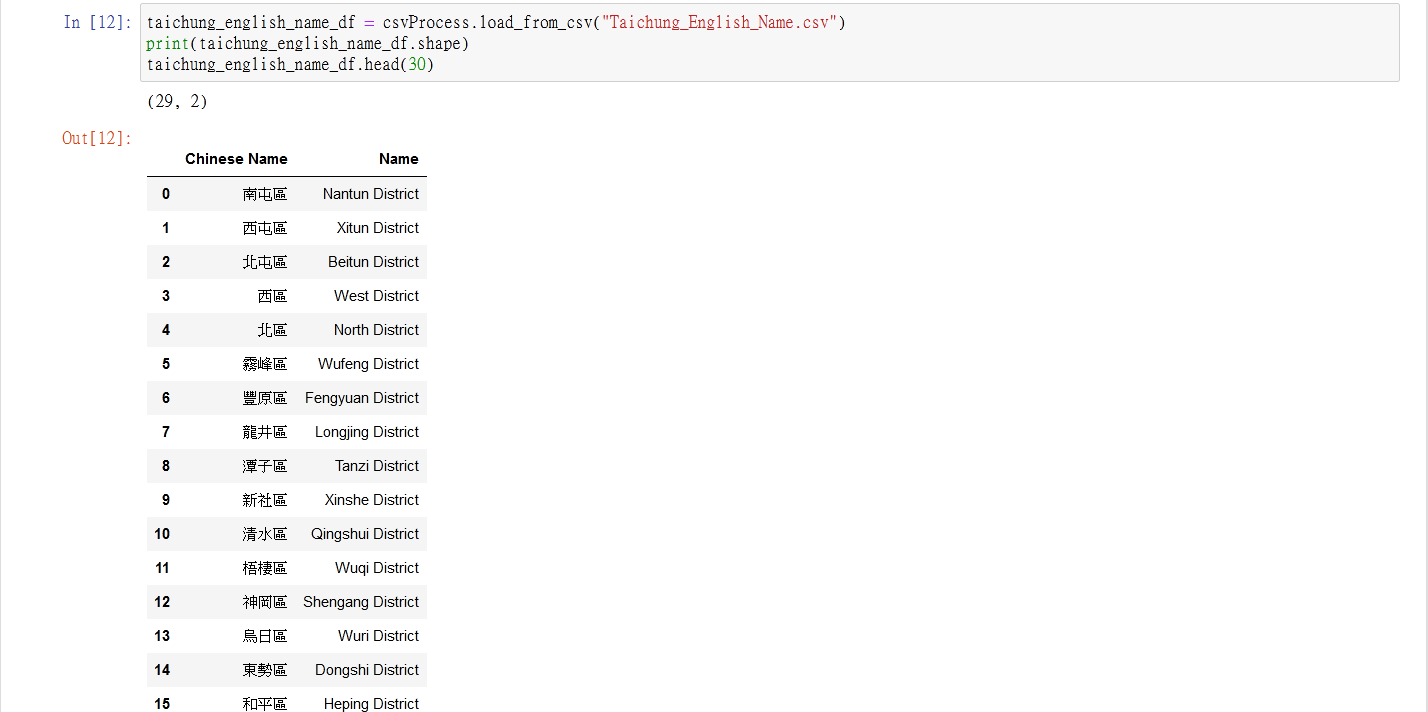
Data source: <https://www.astrocode.net/%E5%8F%B0%E7%81%A3%E5%90%84%E7%B8%A3%E5%B8%82%E5%9C%B0%E5%8D%80%E7%B6%93%E7%B7%AF%E5%BA%A6/>

Description: This is a XML file and we can transfer to JSON file. The file included all **districts** of latitude and longitude in Taiwan.



* 1. English - Chinese glossary of districts in Taiwan:

Data source: <http://gn.moi.gov.tw/geonames/Translation/Translation.aspx>  
Description: Government help to translate the city or districts name between English and Chinese



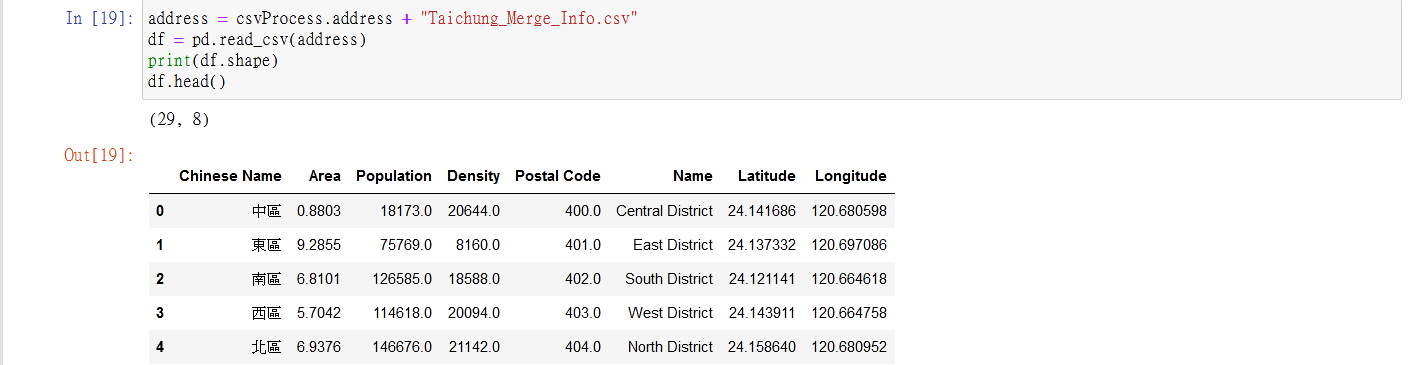


* 1. Venues in each neighborhood of Taichung City:

Data source: Foursquare APIs

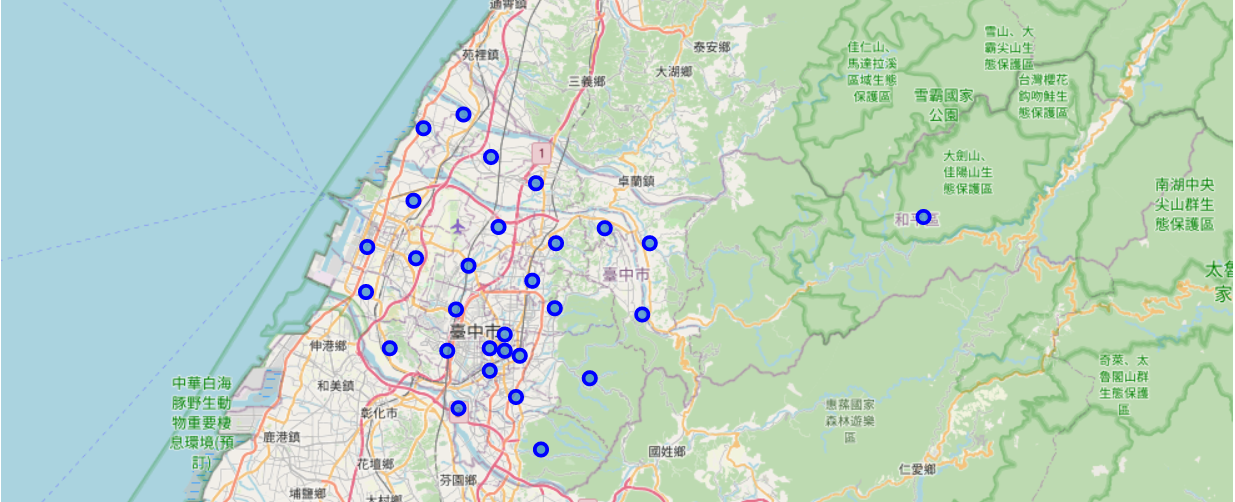
Description: We will get all the venues in each neighborhood by using this API. We can filter these venues to get only restaurants

1. **Visualization and Data Exploration**
   1. Get Merged Dataframe



* 1. Explore the neighborhoods in Taichung (By folium map)

Use geopy library to get the latitude and longitude values of Taichung and create a map of Taichung with neighborhoods superimposed on top.



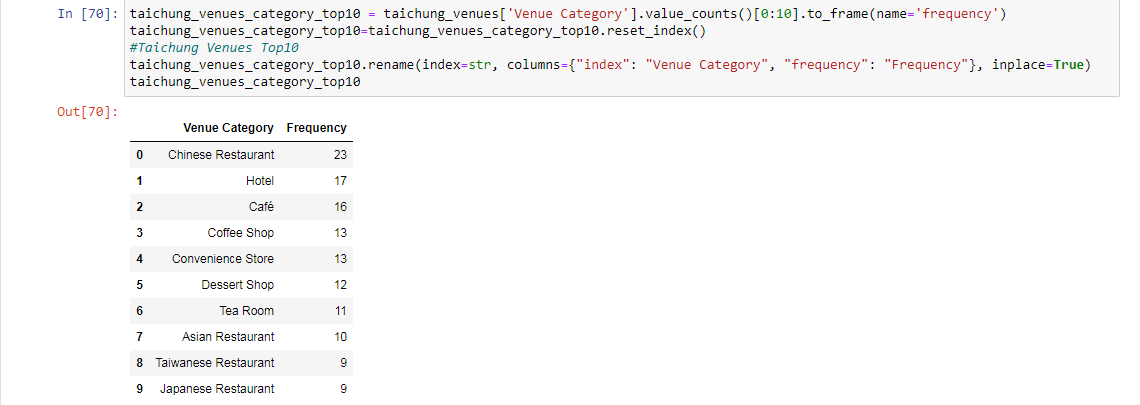
* 1. Explore the neighborhoods in Taichung (By Foursquare API)

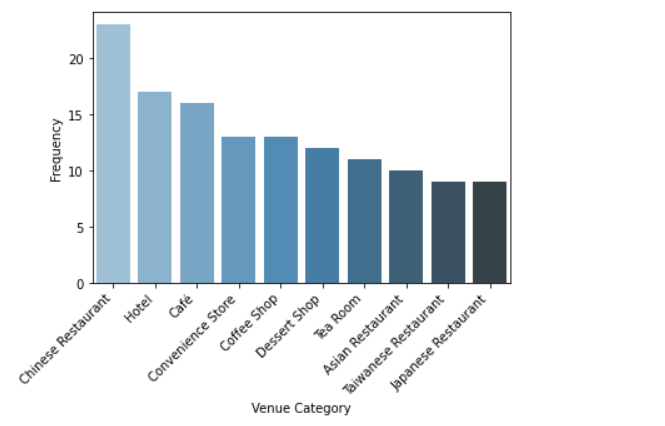
Use Foursquare API and get the top 100 venues that are in Taichung within a radius of 1000 meters.

We found there are total 288 venues in this table.



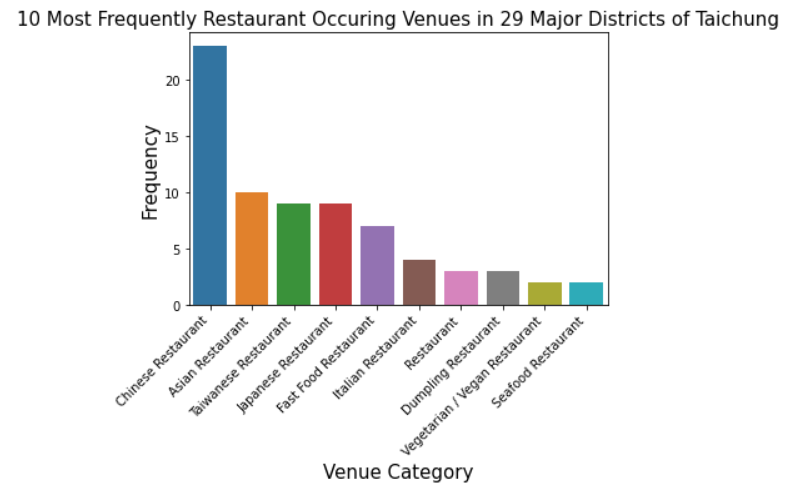
* 1. Visualization the Taichung City dataframe information
     + Rank of Venue Category in Taichung City





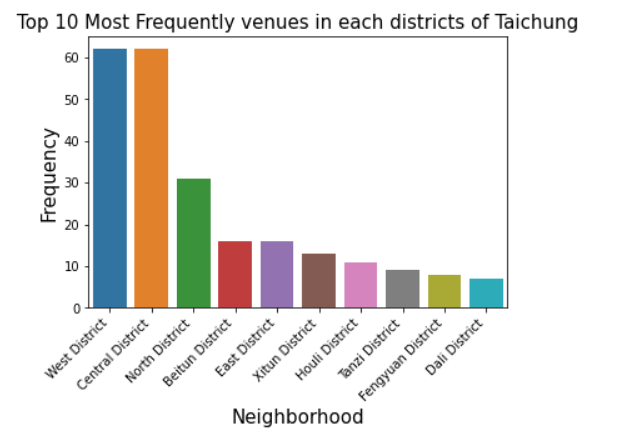
* + - Number of every type of restaurant in Taichung City



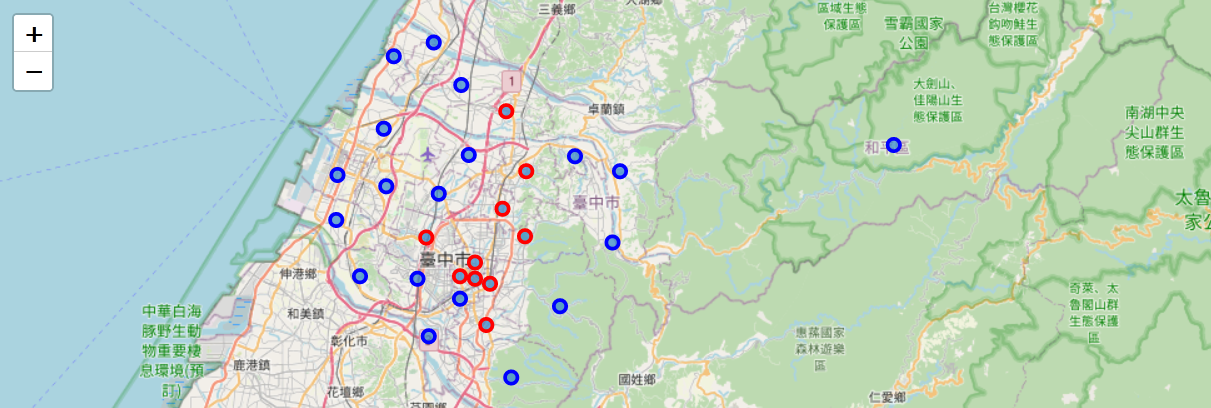


* + - Number of venues in each Neighborhood



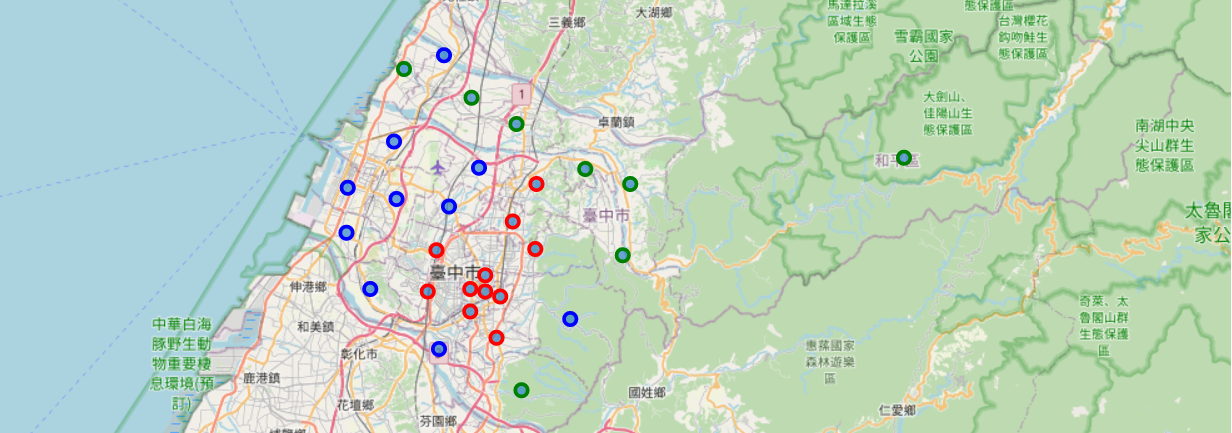


The Folium Map correspond with above bar chart. The map displayed 29 districts in Taichung City. The red points present top 10 districts and the blue points present other 19 districts.



* + - Layer of district density in Taichung city

There are three kind colors in the map. The color red means the density large than 4000, the blue point means the density is from 1000 to 4000 and the green point present the density is small than 1000



* 1. **Results & Discussion**
     + Rank of Venue Category in Taichung City

The bar chart is the top 5 venue category Venue Category in Taichung City, there are restaurant/hotel/cafe/convenience store/dessert ship.

That indicated Taichung is a vary convenience city, you can eat any food and buy anything everywhere. Travers can eat paradise in Taichung.

* + - Number of every type of restaurant in Taichung City

The bar chart is number of every type of restaurant in Taichung city, the top 5 restaurants are Chinese restaurant/Asian restaurant/Taiwanese restaurant/Japanese restaurant /Fast Food restaurant

這表示台中的飲食仍然是傳統華人和台灣人的口味為大宗，但仍受到日本和西方文化的影響，所以日式料理以及速食餐廳也排到了第4和第5名

如果人想來台中開餐廳的話，可以嘗試義式或法式料理或是印度料理…等，這些口味是台中比較少見的，如果做得不錯，會大賣的機率是很高的

* + - Number of venues in each Neighborhood

The bar chart is number of venues in each neighborhood in Taichung city. There are 29 districts in Taichung and here we list top 10 districts West District/Central District/North District/Beitun District/East District/Xitun District/Houli District/Tanzi District/Fengyuan District/Dali District

The map is mapping with bar chart. The red points are top 10 districts with venues number and the blue points are other 19 districts

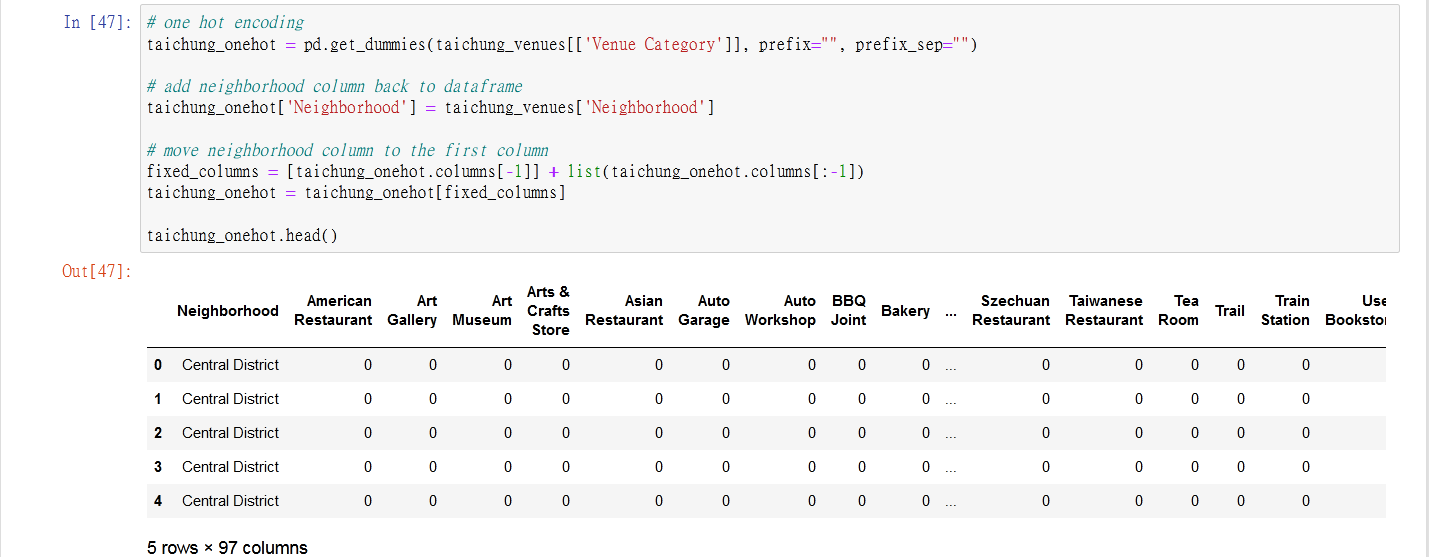
由venues的數量可以看出台中的某些區域會是比較發達與繁榮的，搭配地圖和區域的顏色來看可以更加的了解各區的位置，在top10的區域都是比較靠近市中心，而其他非top10的都是在山區和沿海居多

* + - Layer of district density in Taichung city

此張地圖和上一張的地圖基本上是非常的相似，紅色的區域在2張圖上的位置基本是一樣的，也可以說明了venues的數量和density有一定程度的相關性。

移民者若是考慮方便性以及商業發展的話，可以優先考慮紅色的區域，但若是比較喜歡悠閒或是人口不要這麼稠密的話，可以考慮藍色點的部分，會比較適中，至於綠色的部分就比較靠近山區，這個部分就比較適合旅遊的人來玩玩，不適合久居

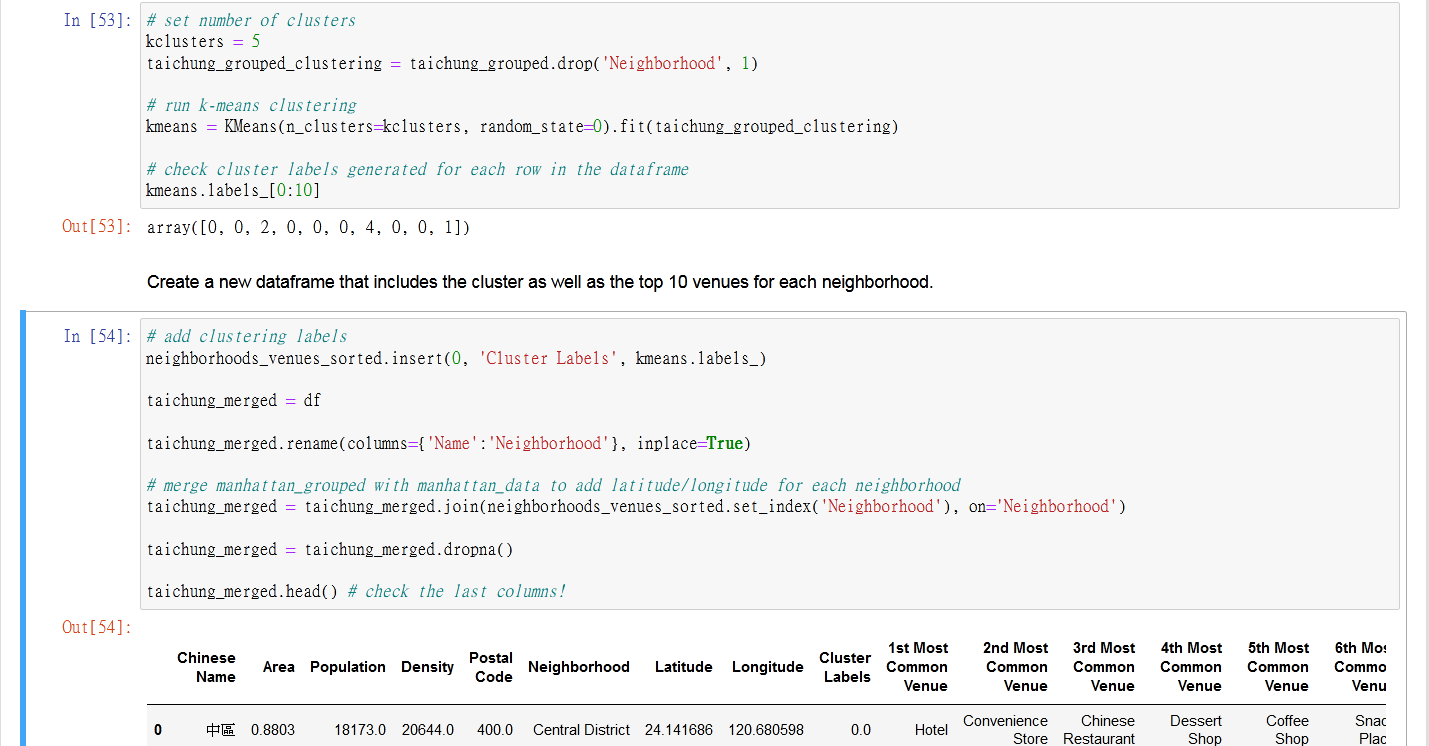
1. **Clustering the neighborhoods**
   1. One hot encoding



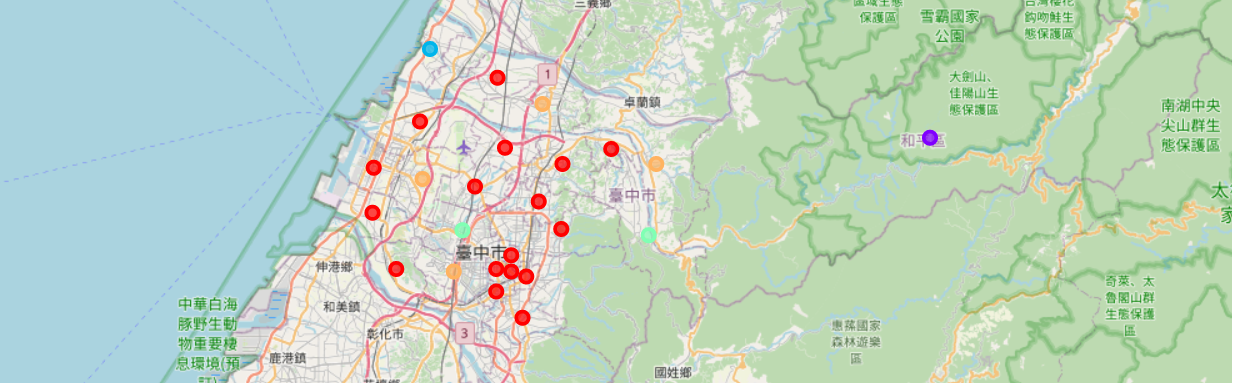
* 1. k-means

Try to cluster these 29 districts based on the venue categories and use K-Means clustering.

We expect the similarities of venue categories will be clustered.



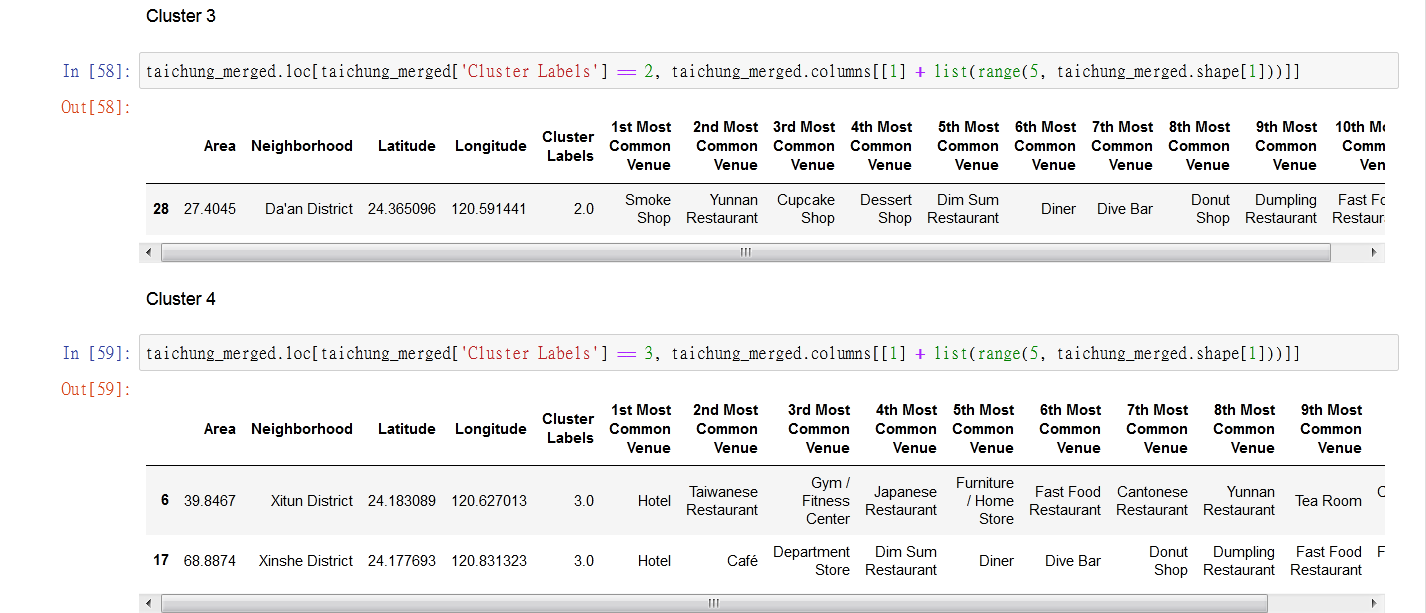
* 1. Visualize the resulting clusters

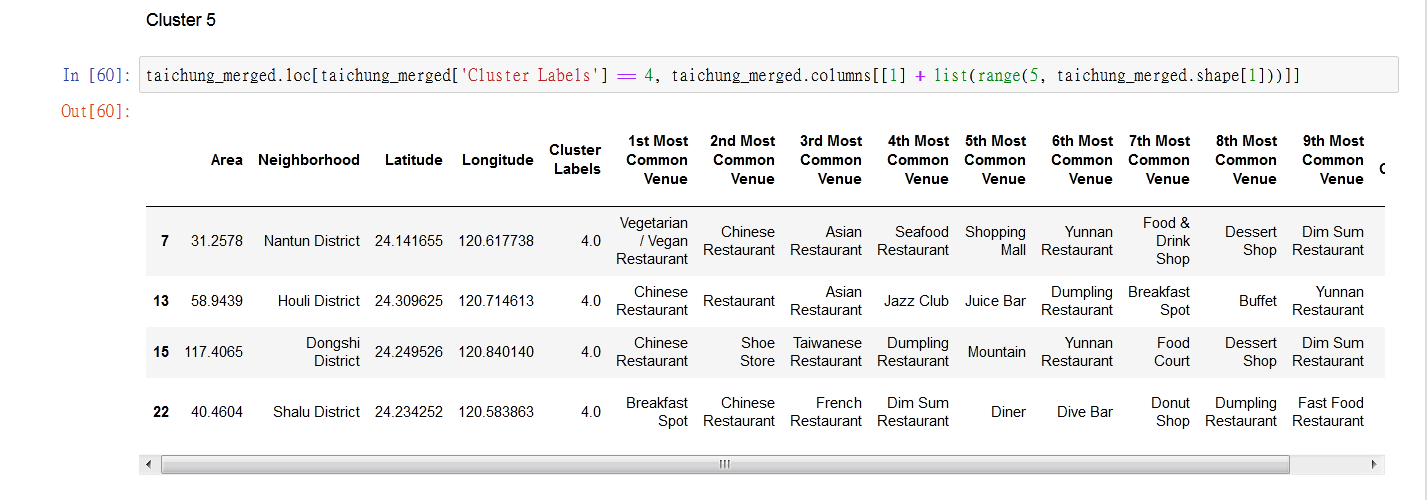


* 1. Examine Clusters









* 1. **Results & Discussion**

Clusters 1 have the greatest number of neighborhoods, cluster 2 and 3 has only one, Cluster 4 has two and Cluster 5 has four.

**Cluster 1** is a group of neighborhoods that has the commonly known for good Hotel, Convenience Store, Café and Coffee Shop

**Cluster 2** recommend Water Park

**Cluster 3** recommend Smoke Shop

**Cluster 4** is main Hotel

**Cluster 5** are almost Chinese Restaurant

從地圖和數據，我們可以初步的了解某些區域的相似性以及差異性。讓人們更可以依據各自的喜好與設定的條件，來選擇自己想居住或是想遊玩的區域

We can understand the similarity and difference on these districts from the map and clustering experiment in first step.

People could base on their favorite or consideration to filter the special districts which they can settle down or visit in Taichung city.

1. **Conclusion**

By using data collected in Foursquare, we have been able to have a small project about analyzing Helsinki venue information. Even though the amount of available data from Foursquare for the Helsinki area is still quite limited, A small guidance for tourists is made in order to provide solutions for future individuals and businesses to improve.

I have made use of some frequently used python libraries to scrap web-data, use Foursquare API to explore the major districts of Tokyo and saw the results of segmentation of districts using Folium leaflet map.

Similarly, data can also be used to solve other problems, which most people face in metropolitan cities. Potential for this kind of analysis in a real-life problem is discussed in great detail. Also, some of the drawbacks and chance for improvements to represent even more realistic pictures are mentioned.

這次的final project結合了許多前面的課程像是收集資料、資料的前處理，網頁爬蟲以及運用Foursquare和clustering的演算法，得出我們想要的資料結果顯示在Map上，讓我們可以初步和了解分析資料，以及可以解決一些問題或是分析商業模式，這讓我覺得很充實，也學到很多東西

這次的project若是再加入房屋的價位以及各區域的工作機會，一定可以更加的精確，得到更多可以分析的資料，來讓移民者或是旅行者更清楚的選擇他們想要的區域

Link to Github: <https://github.com/mtkhanh/Coursera-DataScience/blob/master/FinalPorject.ipynb>