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 is 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 in the world. It was very 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.**

**These characteristics also attract many immigrations come to Taiwan.**

**Taichung is a second big city in Taiwan, its’ urban scale 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 and some are leisure. These districts are so difference!!!**

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

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



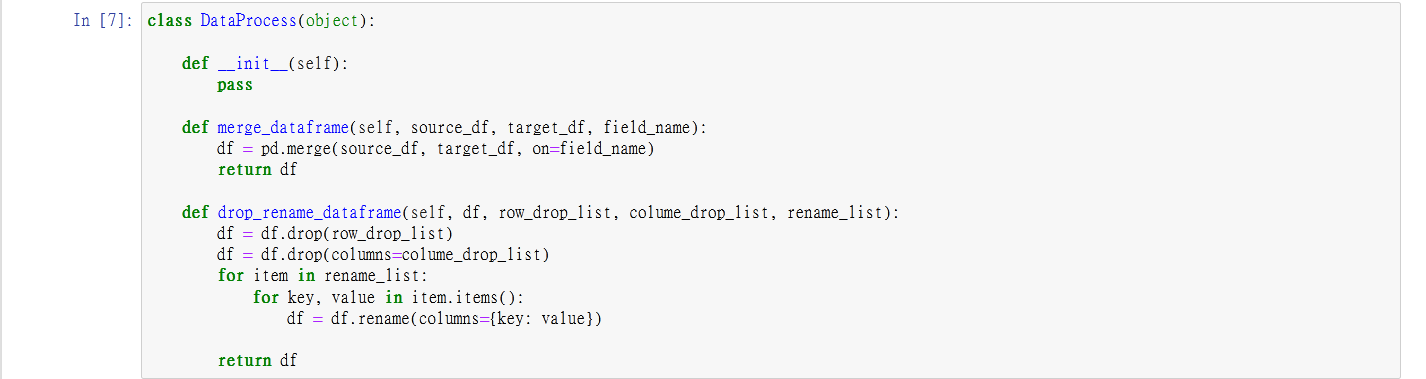
* + 1. Define Json Process Class



* + 1. Define Get Data from Web Class



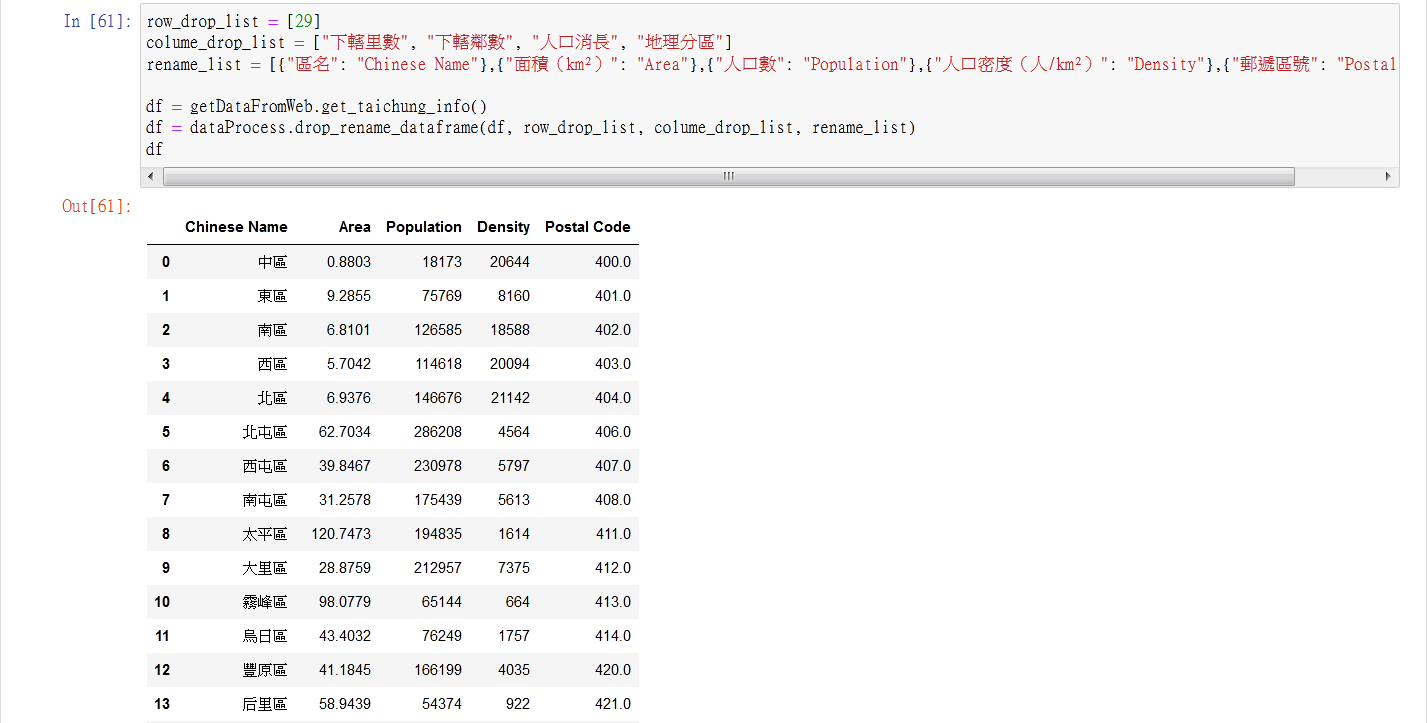
* + 1. Define DataFrame Process Class

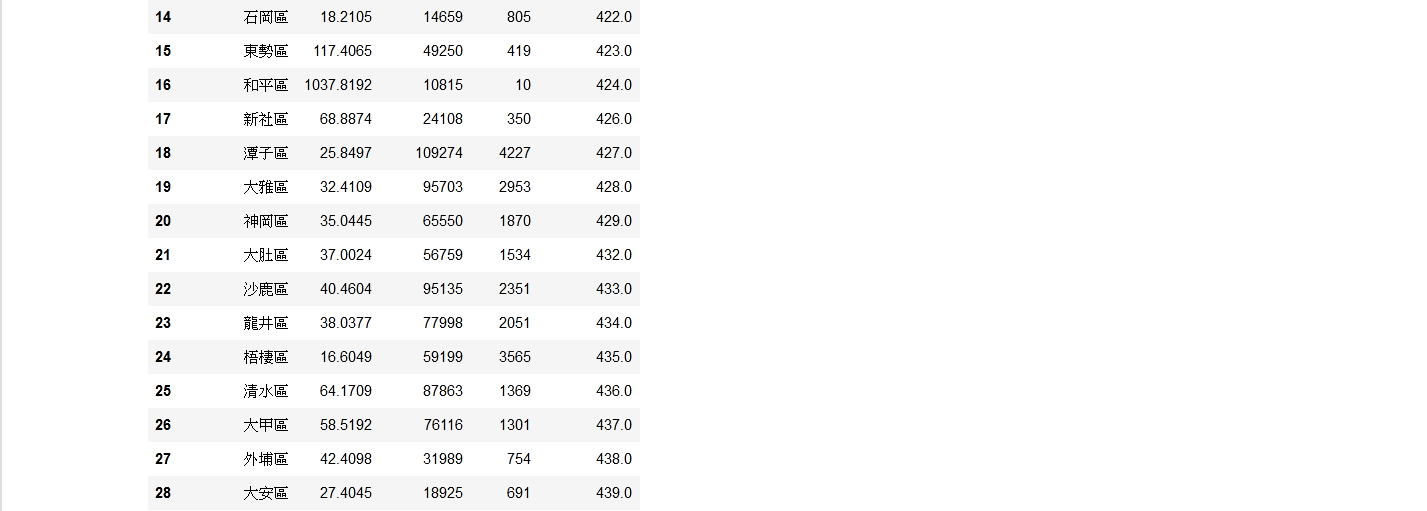


* 1. Taichung wiki:

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

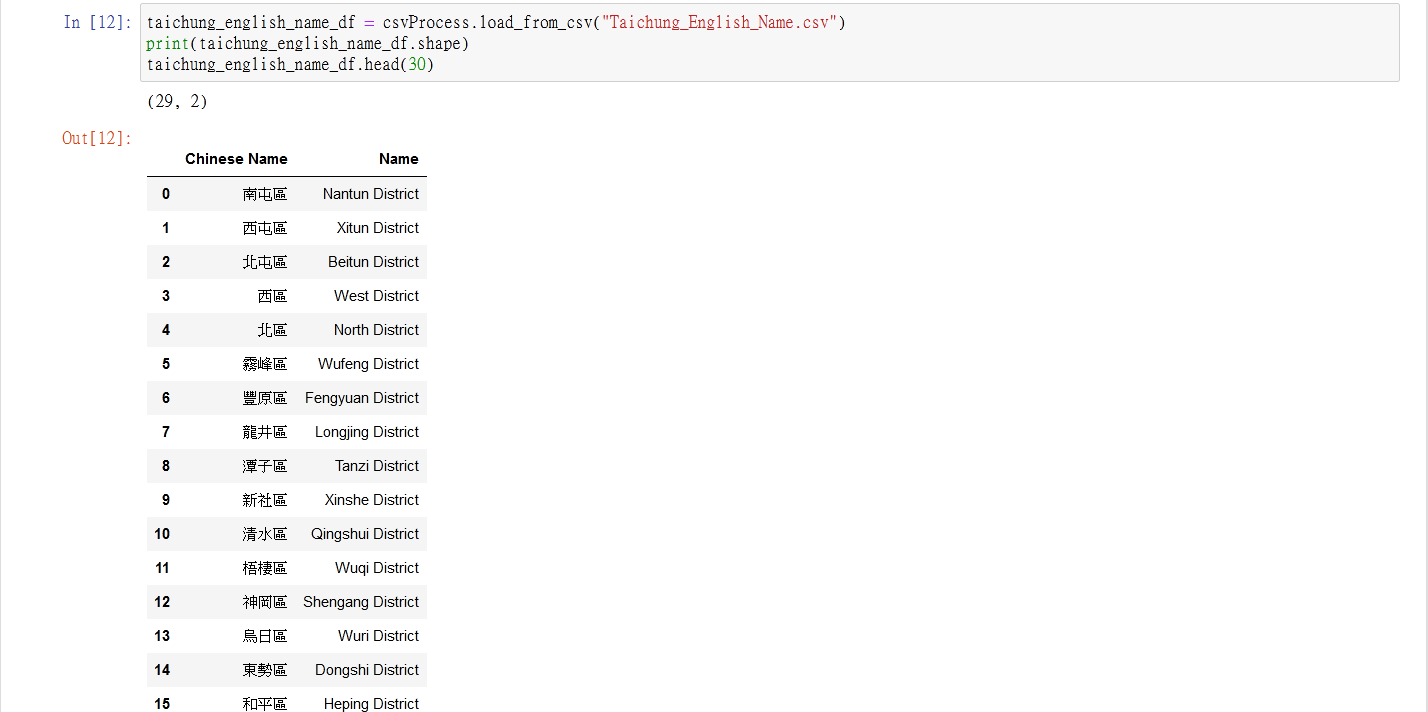
Description: This data set included 29 **districts in Taichung city. There are some** fields we need like: Chinese Name, Area, Density and Postal Code.





* 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. 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. Venues in each neighborhood of Taichung City:

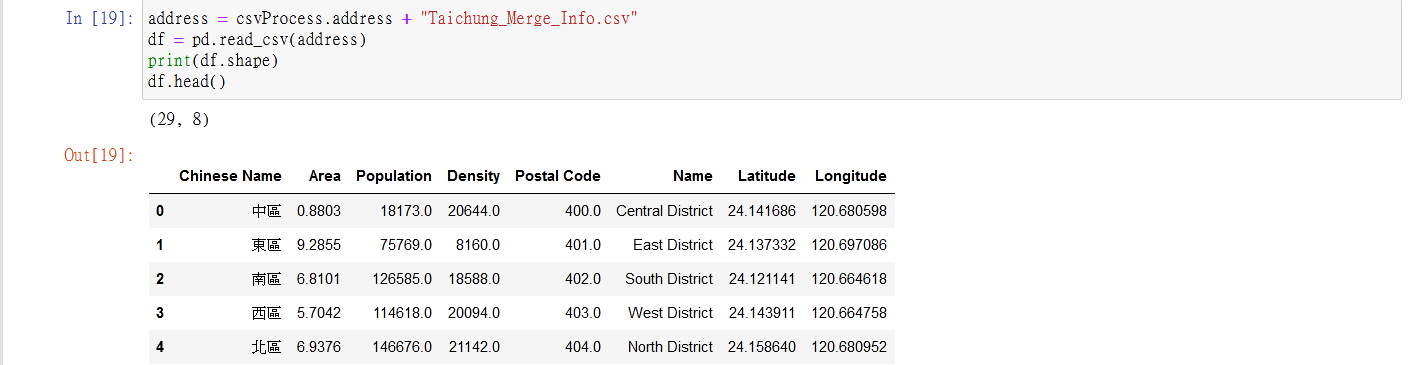
Data source: Foursquare APIs

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

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

By data pre-process and data clean, we will store the data into Taichung\_merge\_Info.csv

Load dataframe from Taichung\_merge\_Info.csv



* 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.

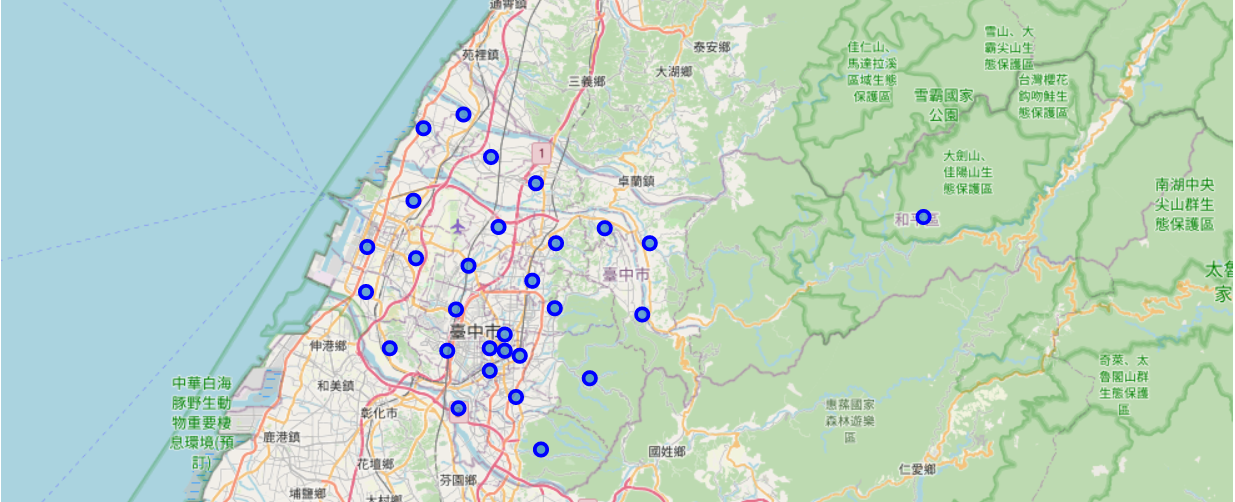


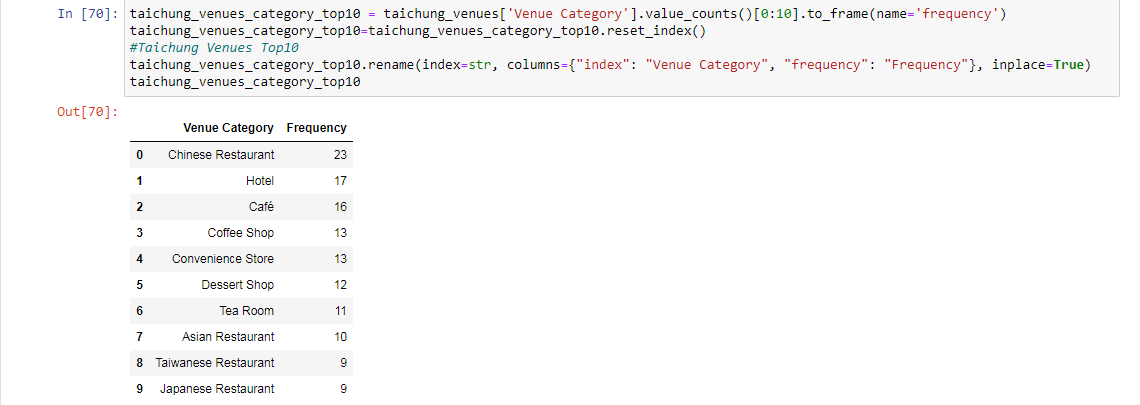
Fig. 1. All districts location in Taichung

* 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
     1. Rank of Venue Category in Taichung City



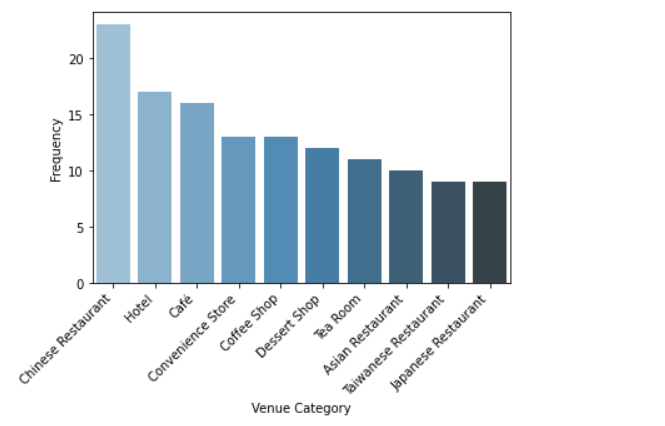


Fig. 2. Top 10 Venue Category in Taichung City

* + 1. Number of every type of restaurant in Taichung City



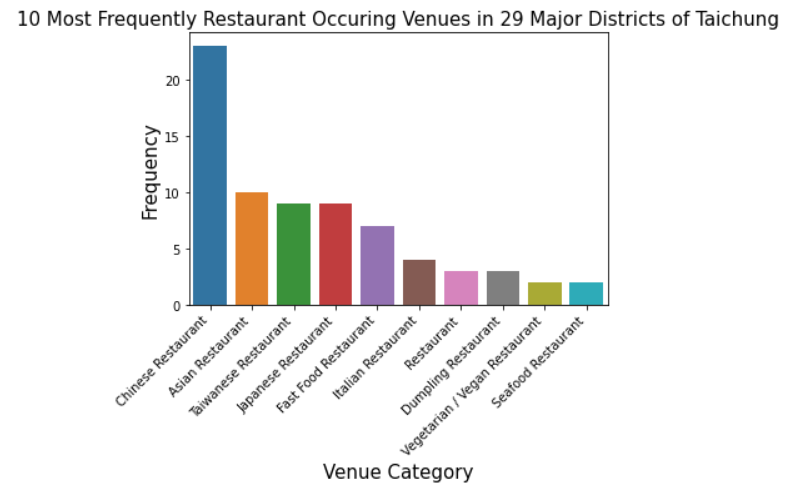


Fig. 3. Top 10 restaurant in Taichung City

* + 1. Number of venues in each Neighborhood



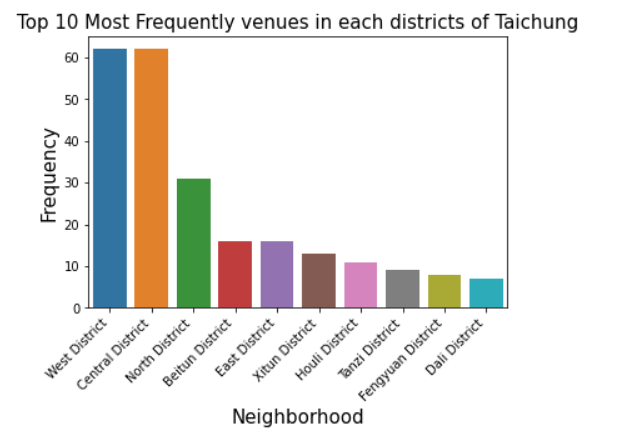


Fig. 4. Top 10 most frequently venues in each districts of Taichung City

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.

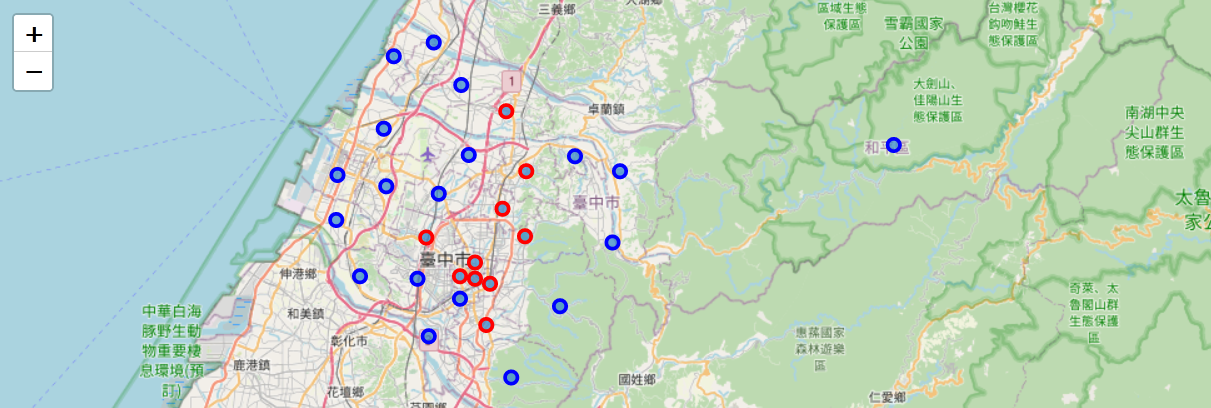


Fig. 5. Top 10 most frequently venues in each districts locations

* + 1. Layer of district density in Taichung city

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

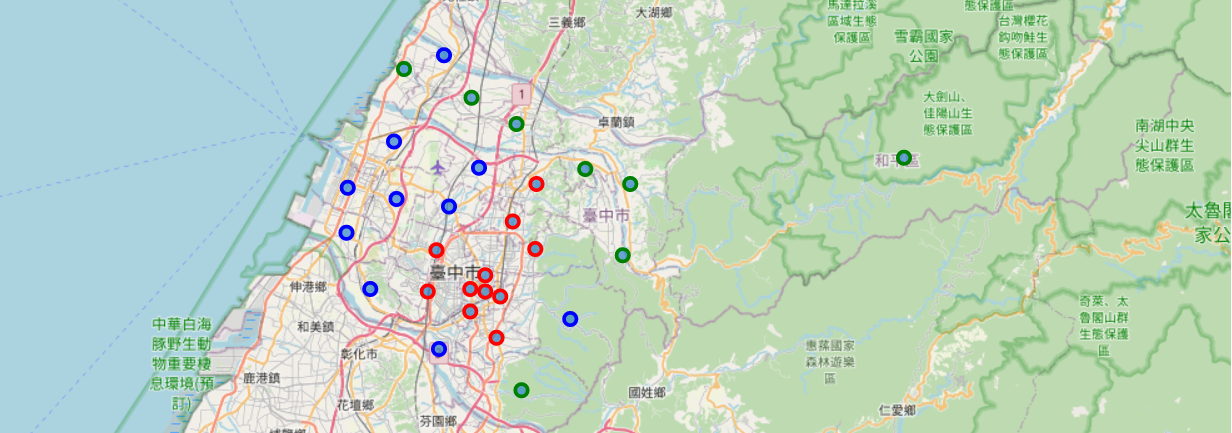


Fig. 6. Layer of district density locations in Taichung city

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

The bar chart (Fig. 2.) is the top 5 venue category Venue Category in Taichung City, there are restaurant/hotel/cafe/convenience store/dessert ship.

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

* + 1. Number of every type of restaurant in Taichung City

The bar chart (Fig. 3.) 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

We can get some analysis from this bar chart.

**The First**: Chinese restaurant/Asian restaurant/Taiwanese restaurant are top 3, that percent the food culture still prefer traditional Chinese favor and Taiwanese local favor in Taichung.

**The Second**: Japanese restaurant /Fast Food restaurant are fourth and fifth, that percent the food culture was impact by Japan and Western culture (USA) in recent twenty years.

That means there are more and more food favors in Taichung. If someone want to create restaurant and I will suggest Italian or French cuisine or Indian food. These favors are rare in Taichung city and these restaurants will be popular if their meals vary are delicious or special.

* + 1. Number of venues in each Neighborhood

The bar chart (Fig. 4) 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 and Dali District

The map (Fig. 5.) is mapping with bar chart. The red points are top 10 districts with venues number and the blue points are other 19 districts

We can get some analysis from this bar chart. Some districts are more urbanization and prosperity and some place are still rural and original.

We also get other analysis from the map. The top 10 are concentrated on the central of Taichung and others are located on mountain or coastal.

* + 1. Layer of district density in Taichung city

This map is corresponding with the map of “Number of venues in each Neighborhood”.

This map (Fig. 6.) is similar with previous map (Fig. 5.). The red points located in the same districts in these two maps. That means the number of venues and density with high relationship.

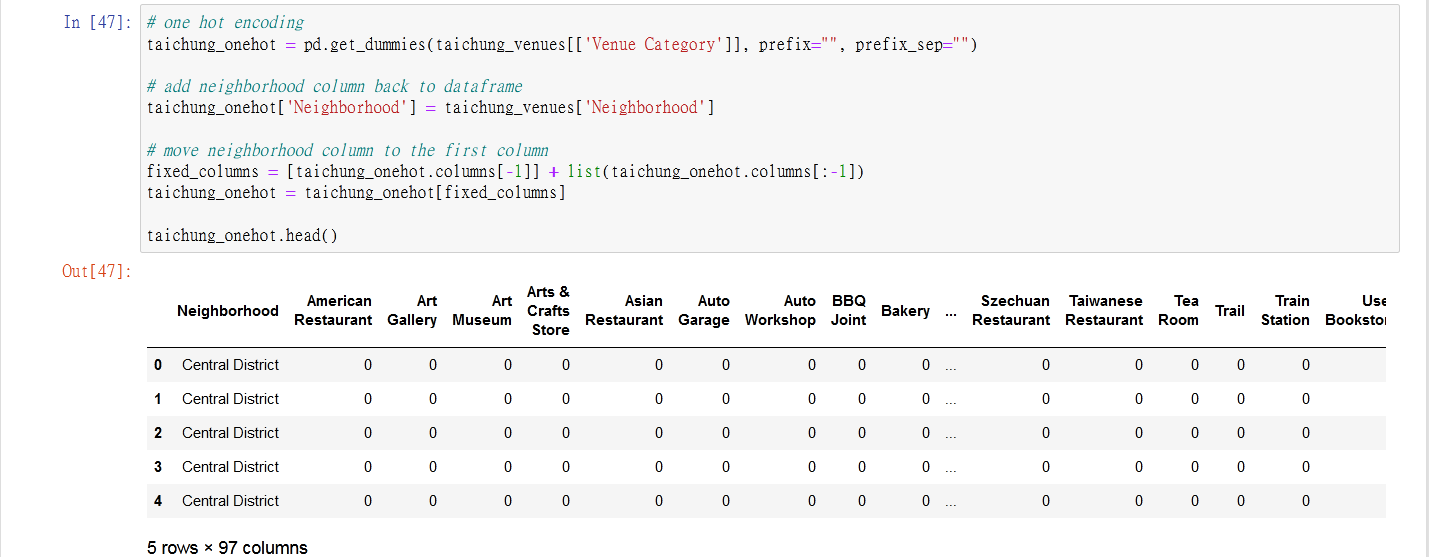
The property of red point is convenience and business development but with large population and noisy in the night.

The property of green point is quiet and with natural environment and small population but is not convenience.

The property of blue point is between red point and green point. With leisure and moderate population.

This analysis can recommend **immigrations or travers which** districts they want to journey or settle down.

1. **Analyze Each Neighborhood and Clustering the neighborhoods**
   1. Analyze Each Neighborhood
      1. One hot encoding

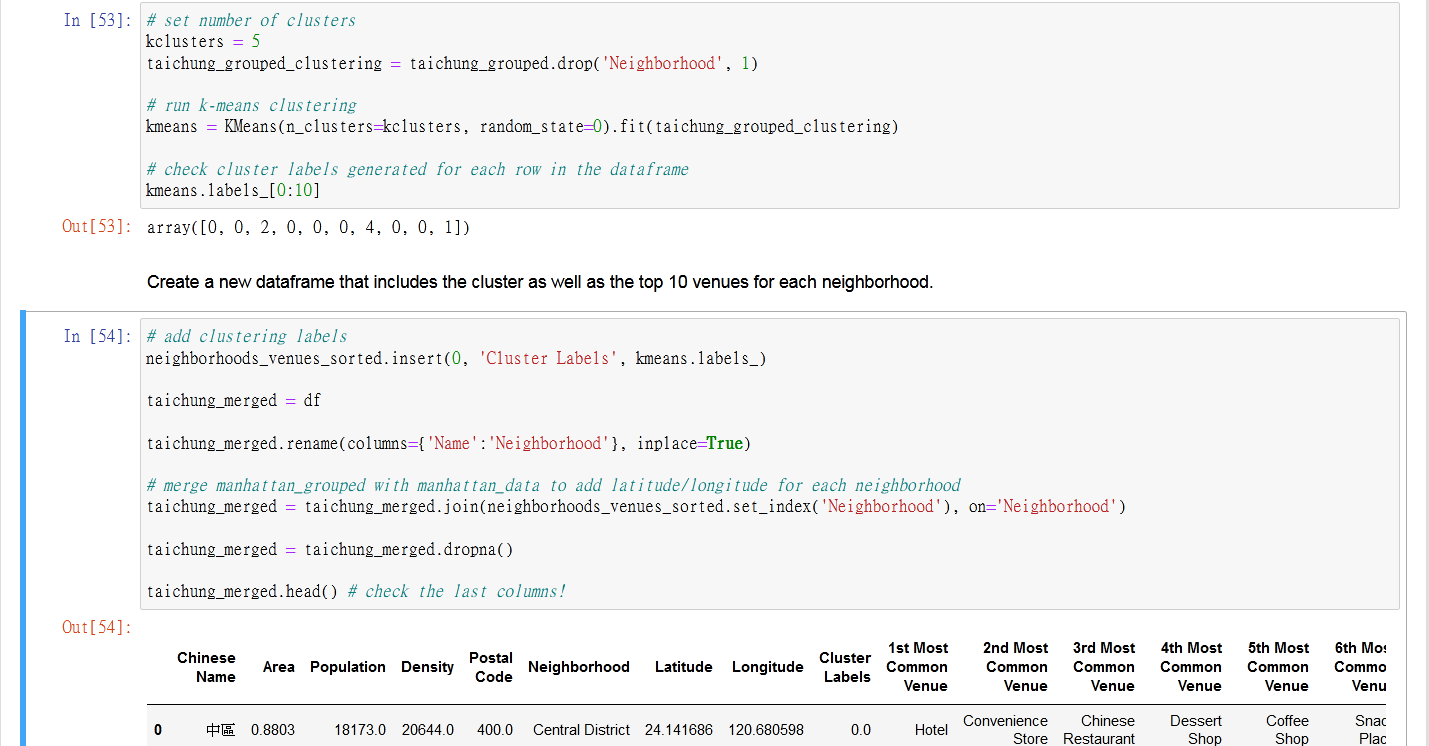


* + 1. Analyze top 10 venues



* 1. Clustering the neighborhood
     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

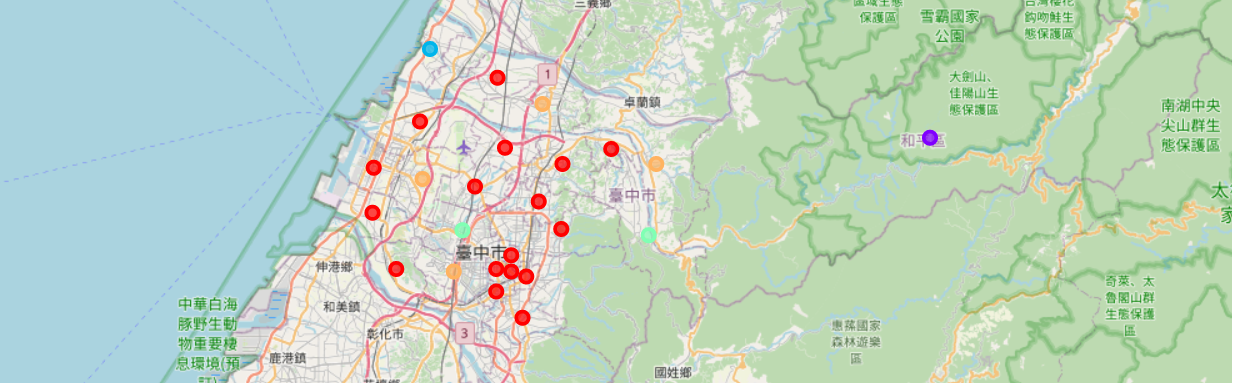
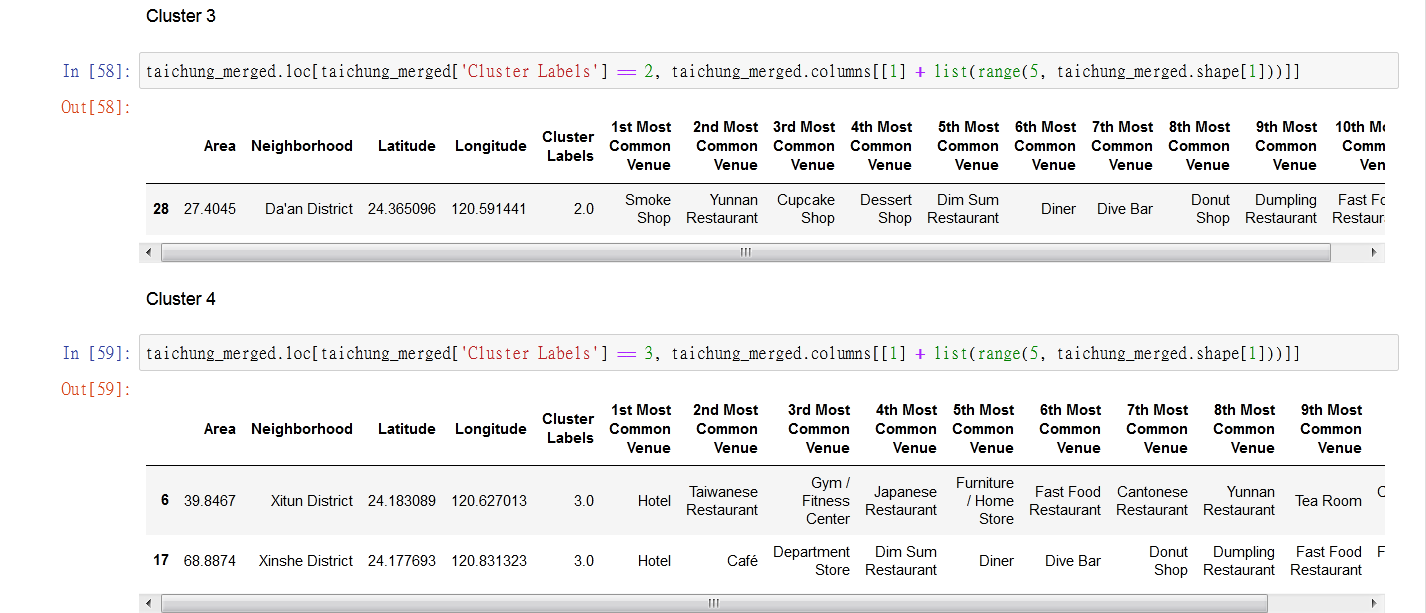


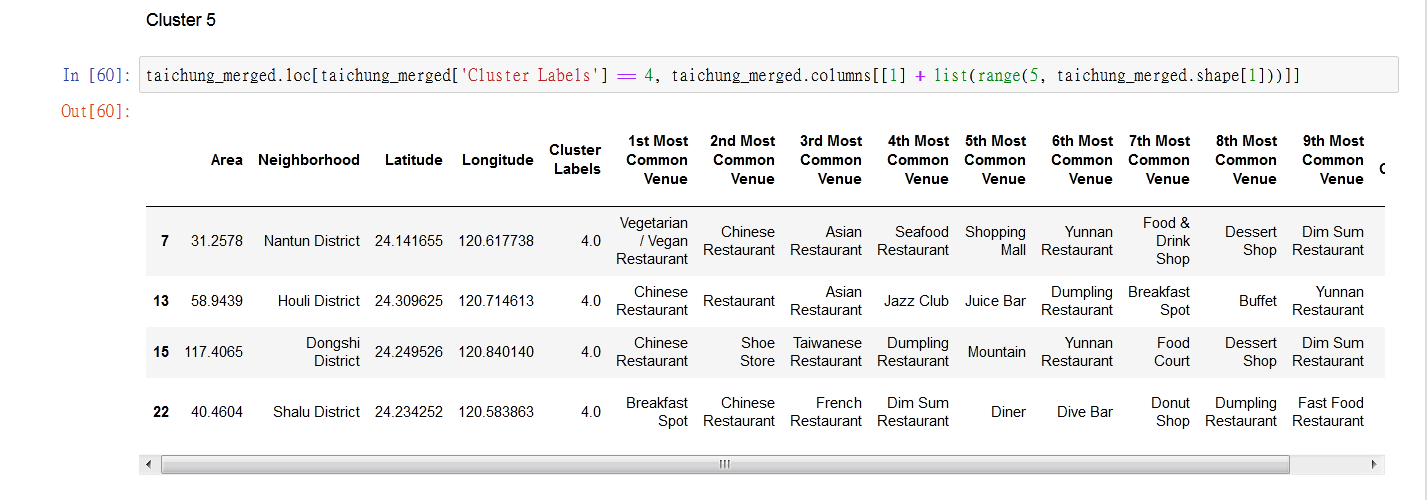
Fig. 7. The districts locations of the clustering result

* 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

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**

We used the previous skills and technology like: Data collection, Date pre-process, Data clean, web crawler, Foursquare API, Supervisor machine learning, Unsupervised learning, Folium Map in this final project.

By these skills and knowledge, we can analysis data and recommend some suggestions to solve problem or more understand business mode. This project let me learned a lot of machine learning technology and understand what is data science and how to do and complete it.

In this final project, we can add house price and the job opportunity in each district that can raise the data integrity. Use the more detail data, we can get clearly analysis report and improved the recommendation to tell **immigrations and travers how to choose the** districts more clearly.

Link to Github: https://github.com/prophetstorbrianchen/Coursera\_Capstone/blob/main/week5/Capstone%20Project%20-%20The%20Battle%20of%20Neighborhoods%20-%20Taichung%20vision.ipynb