**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 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



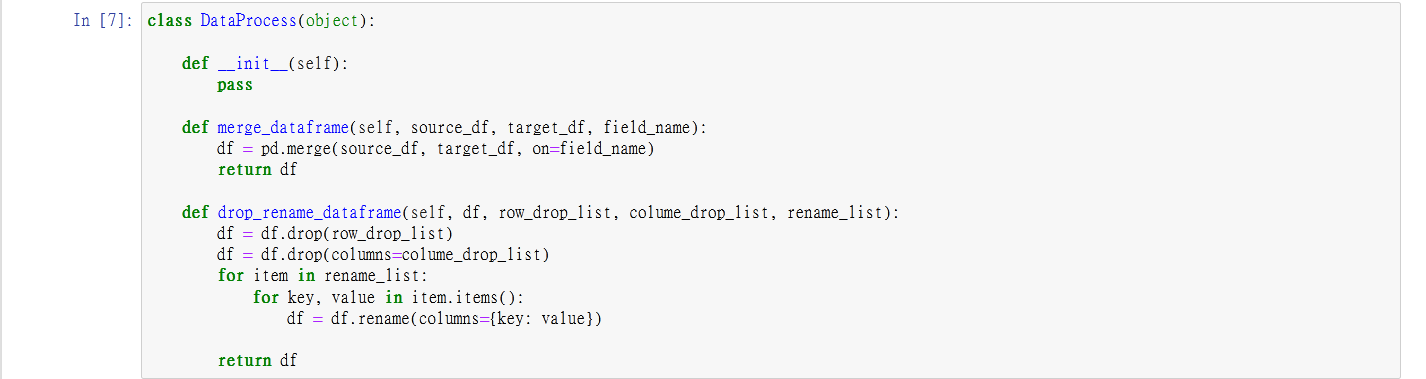
* + - Define Json Process Class



* + - Define Get Data from Web Class



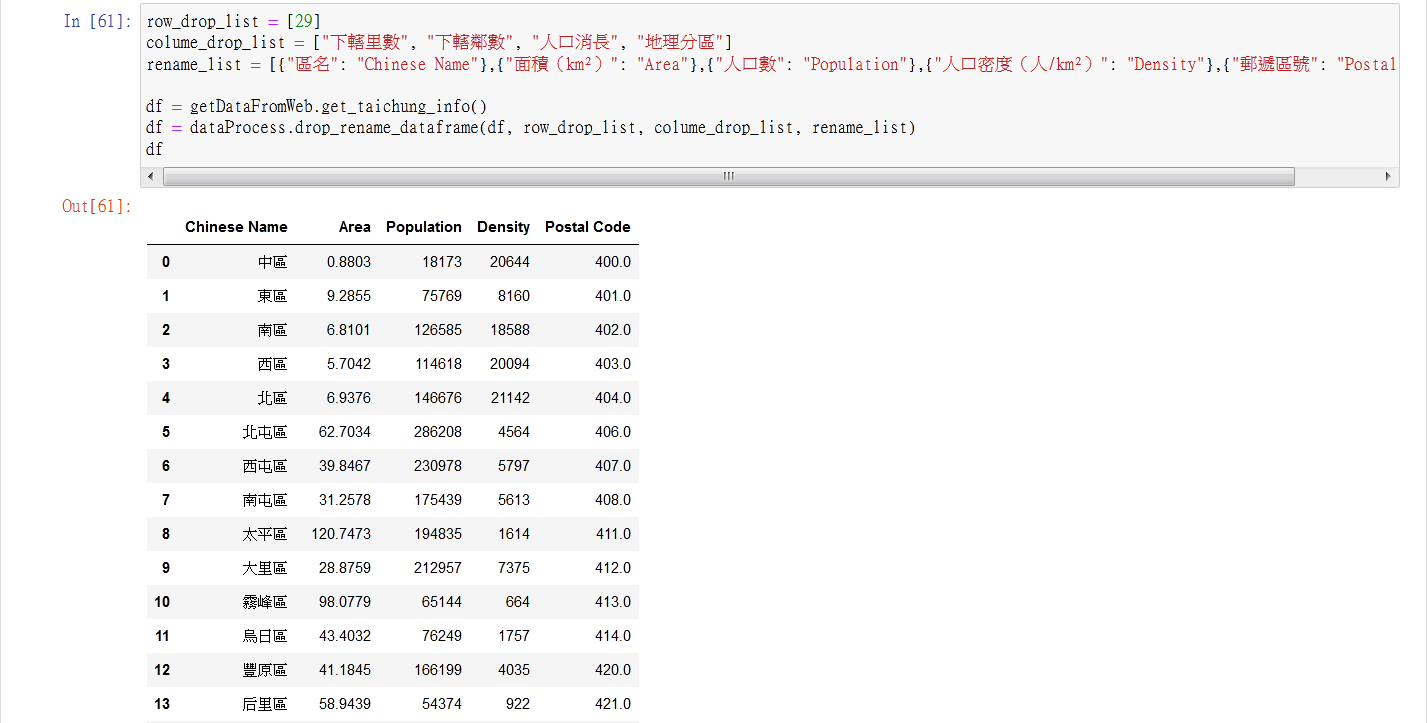
* + - Define DataFrame Process Class

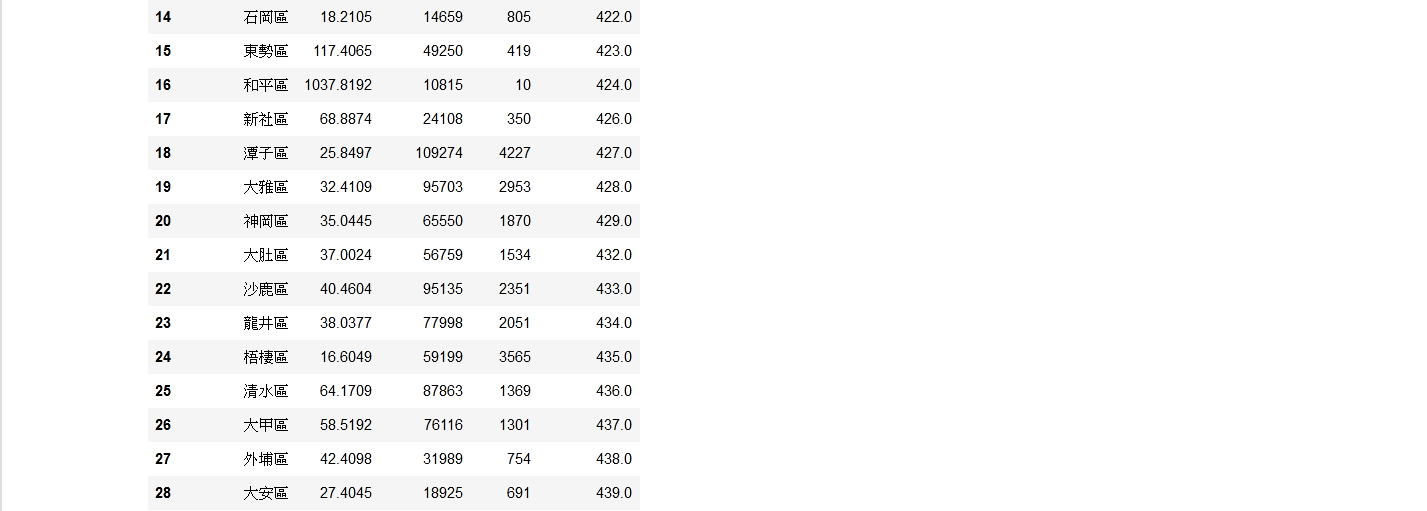


* 1. Taichung wiki:

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:

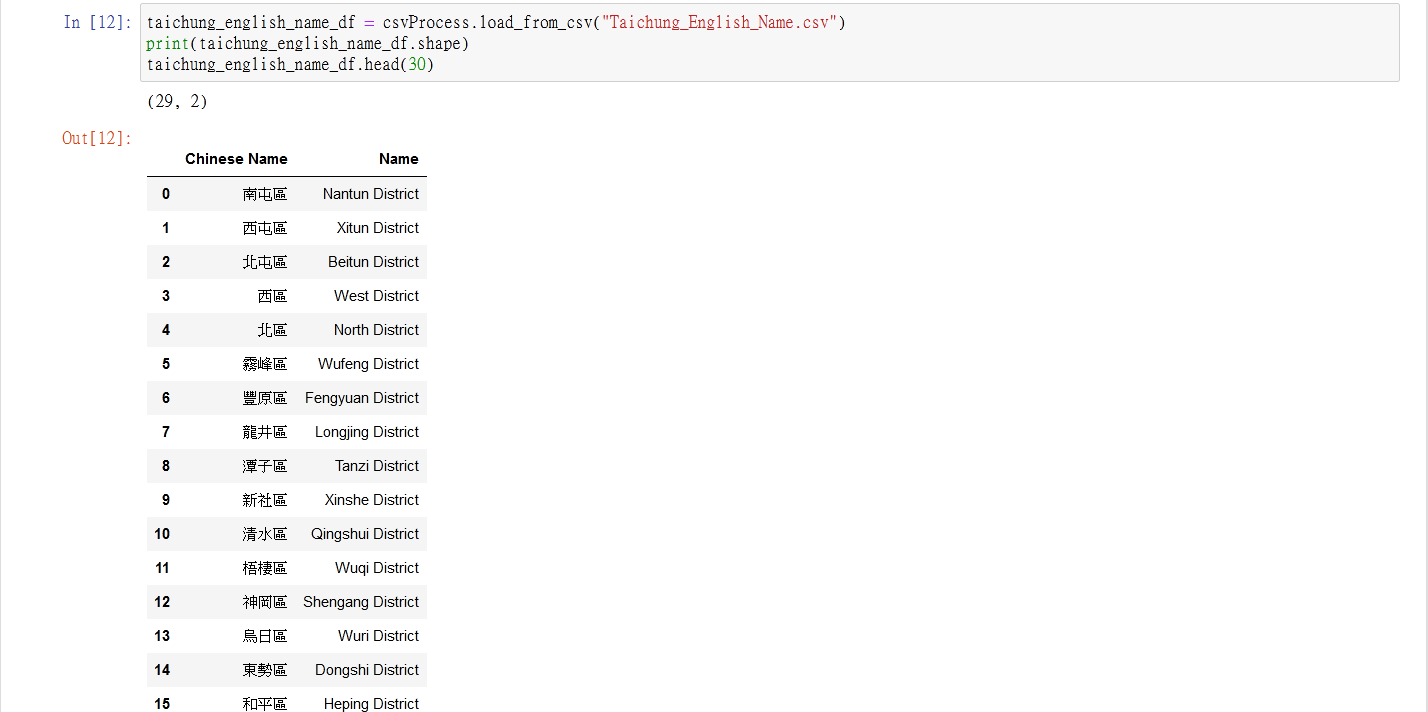
Data source: <https://www.astrocode.net/%E5%8F%B0%E7%81%A3%E5%90%84%E7%B8%A3%E5%B8%82%E5%9C%B0%E5%8D%80%E7%B6%93%E7%B7%AF%E5%BA%A6/>

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: 台灣政府的一個網站,專門為台灣的中文地區轉成英文名字



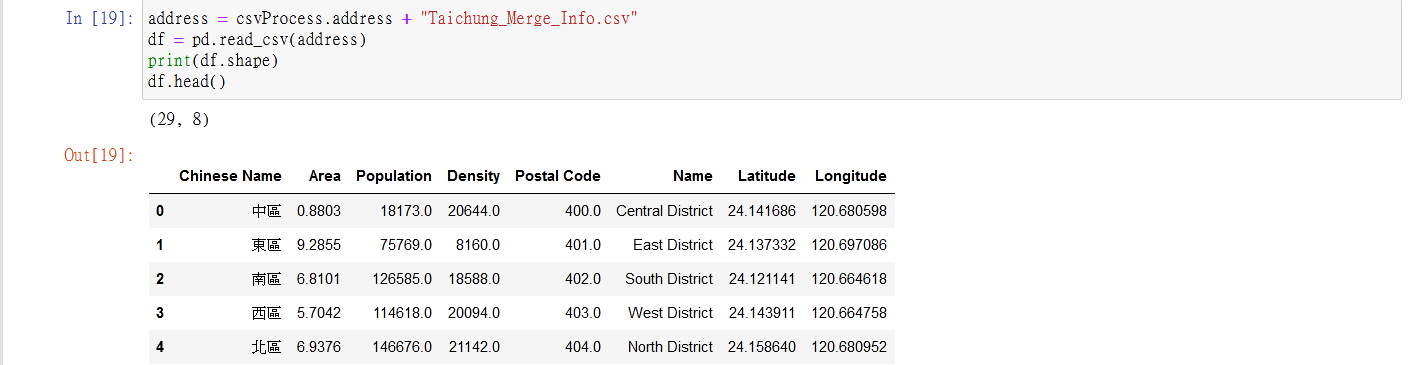


* 1. Venues in each neighborhood of Taichung City:

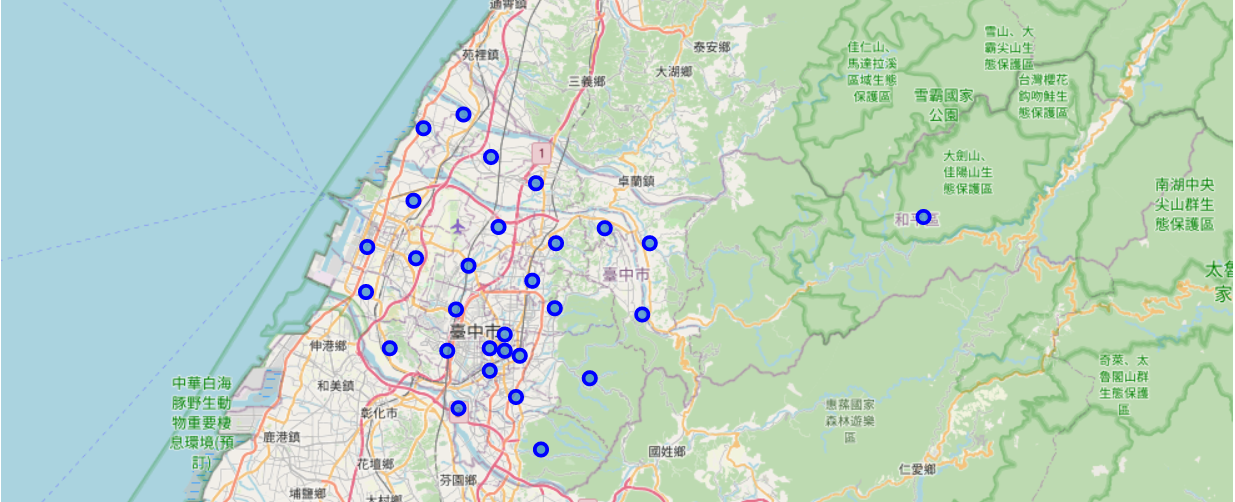
Data source: Foursquare APIs

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



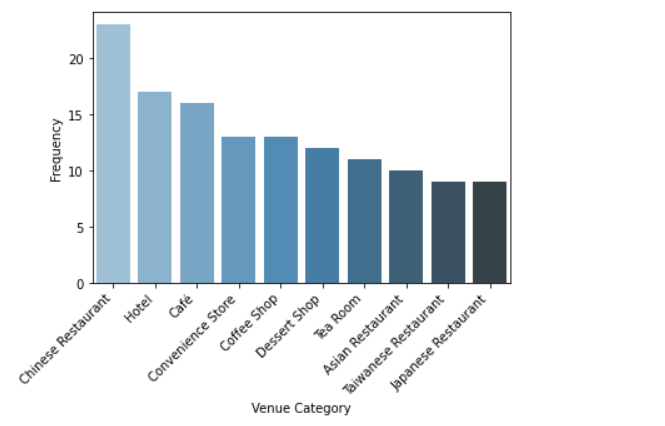
* 1. Explore the neighborhoods in Taichung (By folium map)



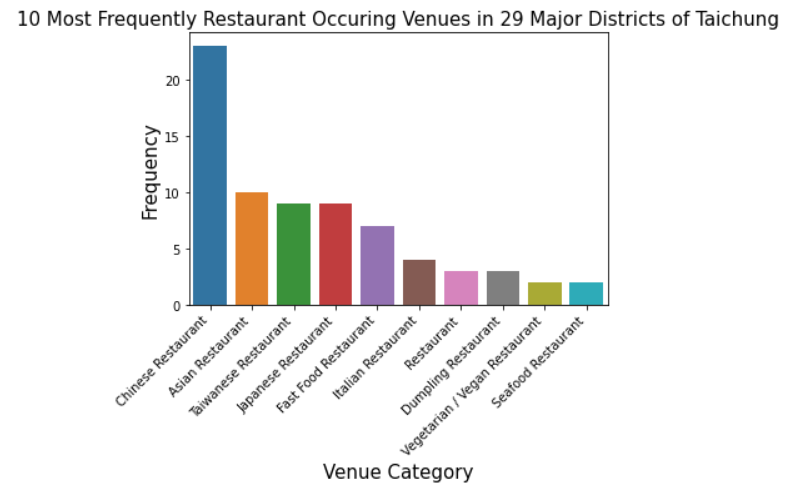
* 1. Explore the neighborhoods in Taichung (By Foursquare API)



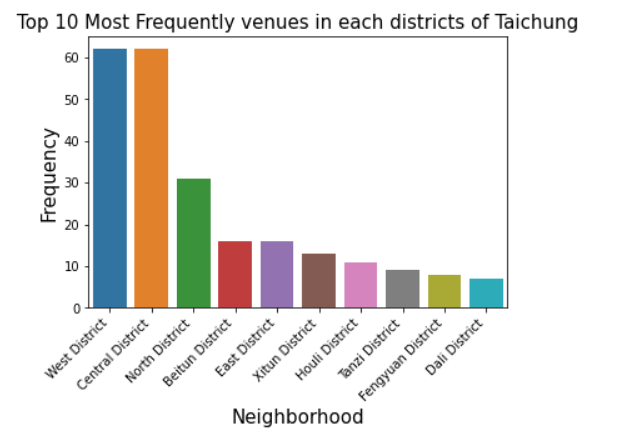
* 1. Visualization the Taichung City dataframe information
     + Rank of Venue Category in Taichung City



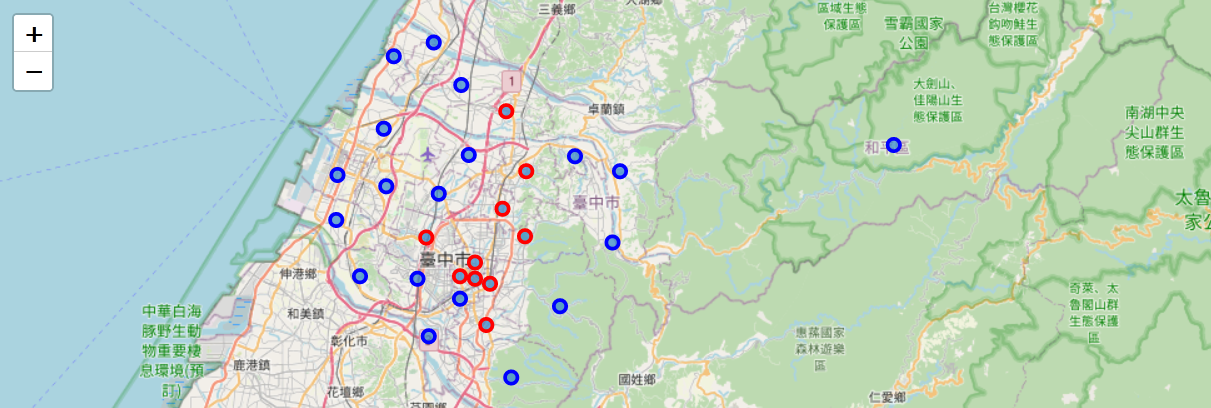
* + - Number of every type of restaurant in Taichung City



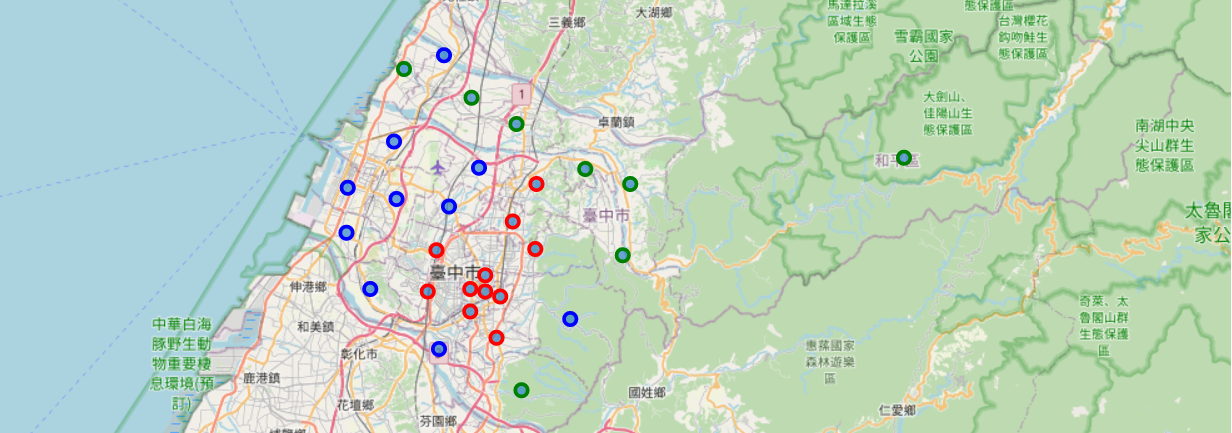
* + - Number of venues in each Neighborhood



對照Map

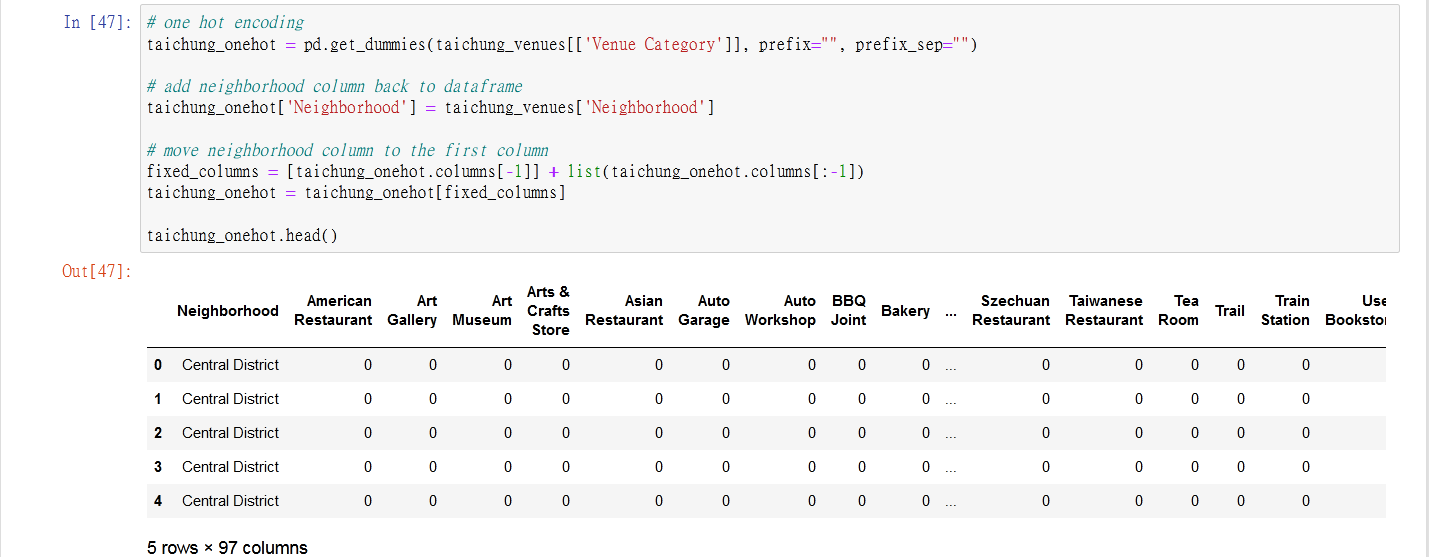


* + - Layer of district density in Taichung city

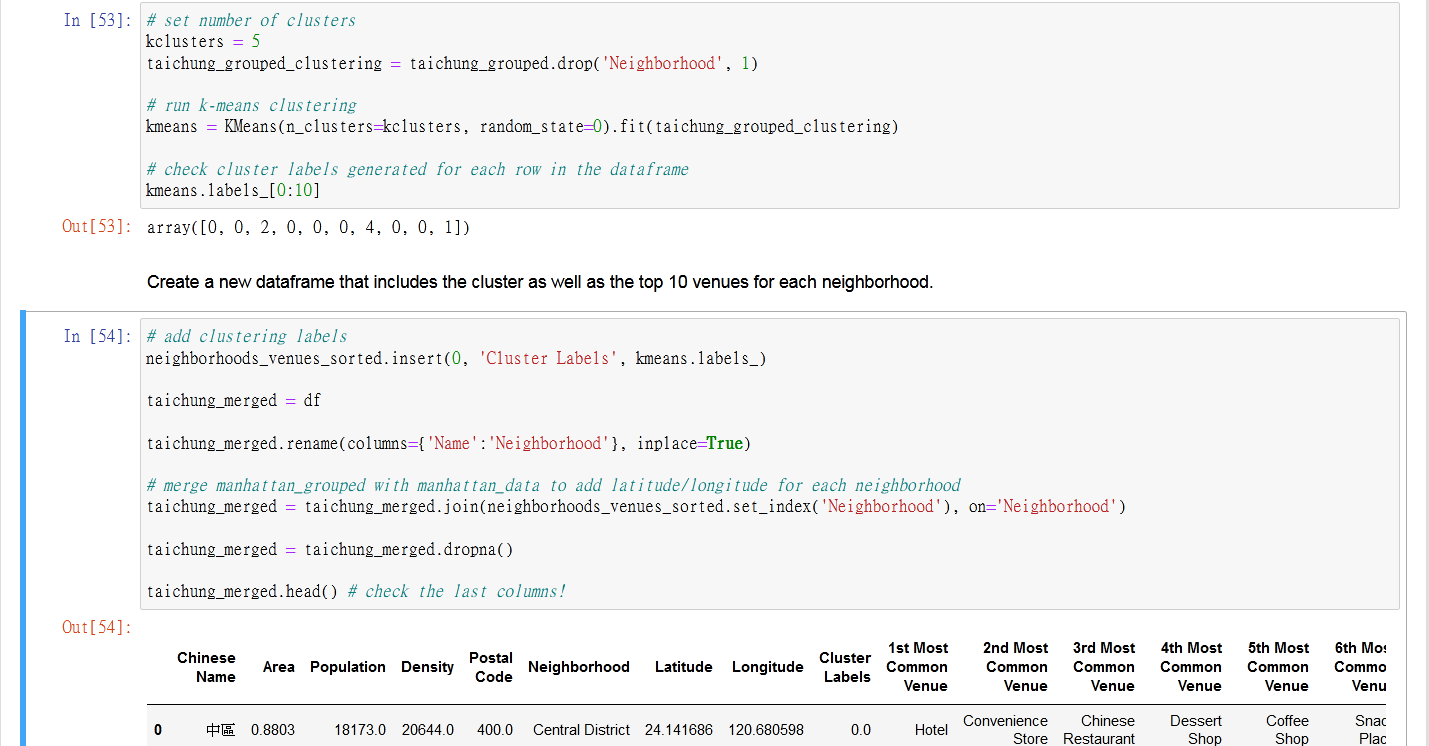


* 1. **Results & Discussion**
     + Rank of Venue Category in Taichung City
     + Number of every type of restaurant in Taichung City
     + Number of venues in each Neighborhood
     + Layer of district density in Taichung city

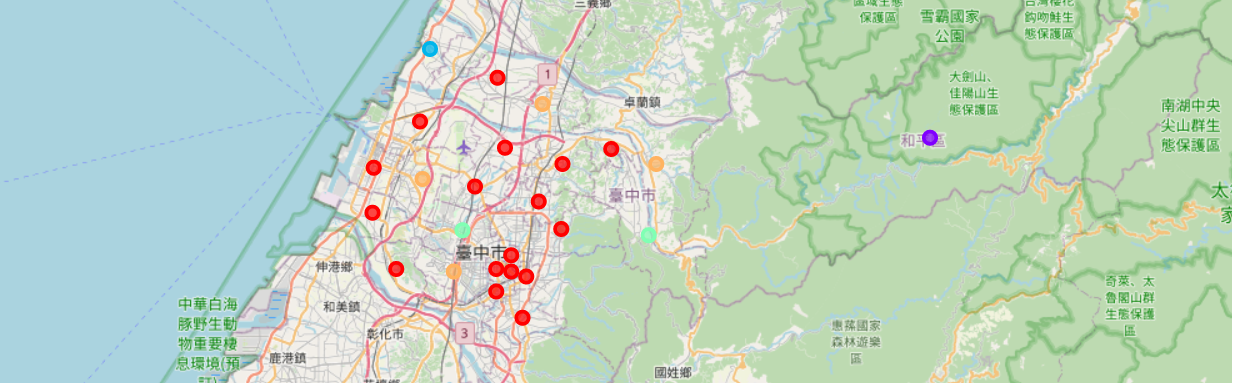
1. **Clustering the neighborhoods**
   1. one hot encoding



* 1. k-means



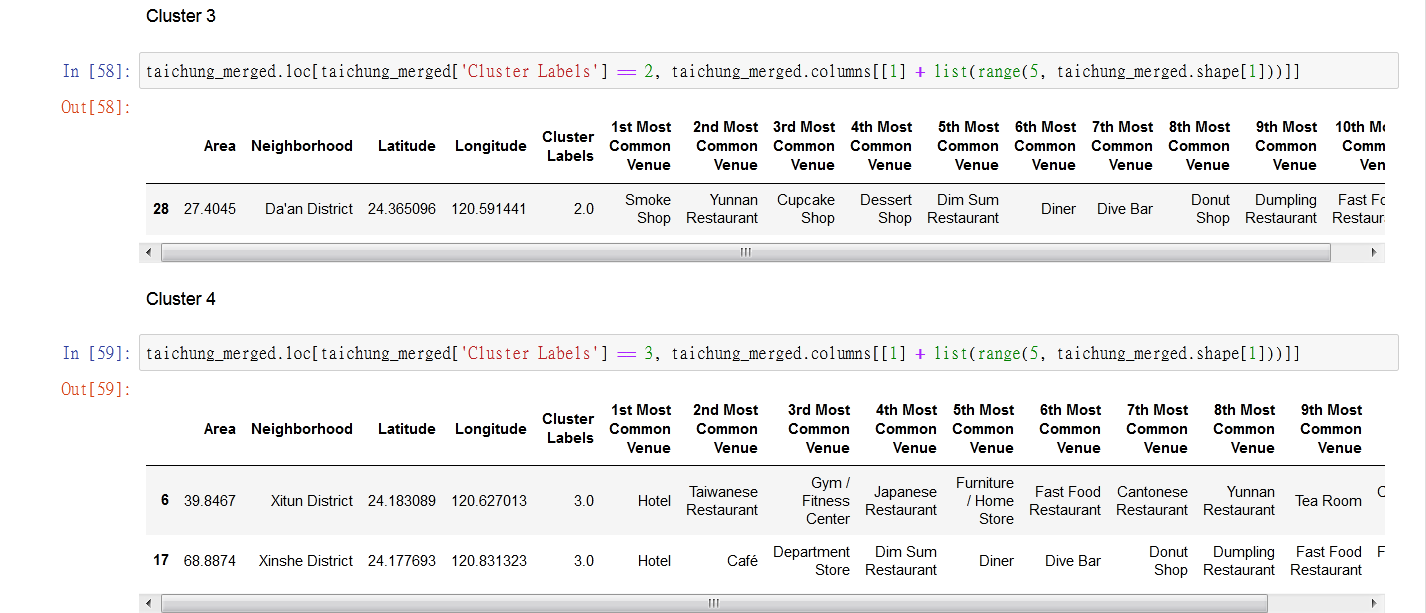
* 1. visualize the resulting clusters

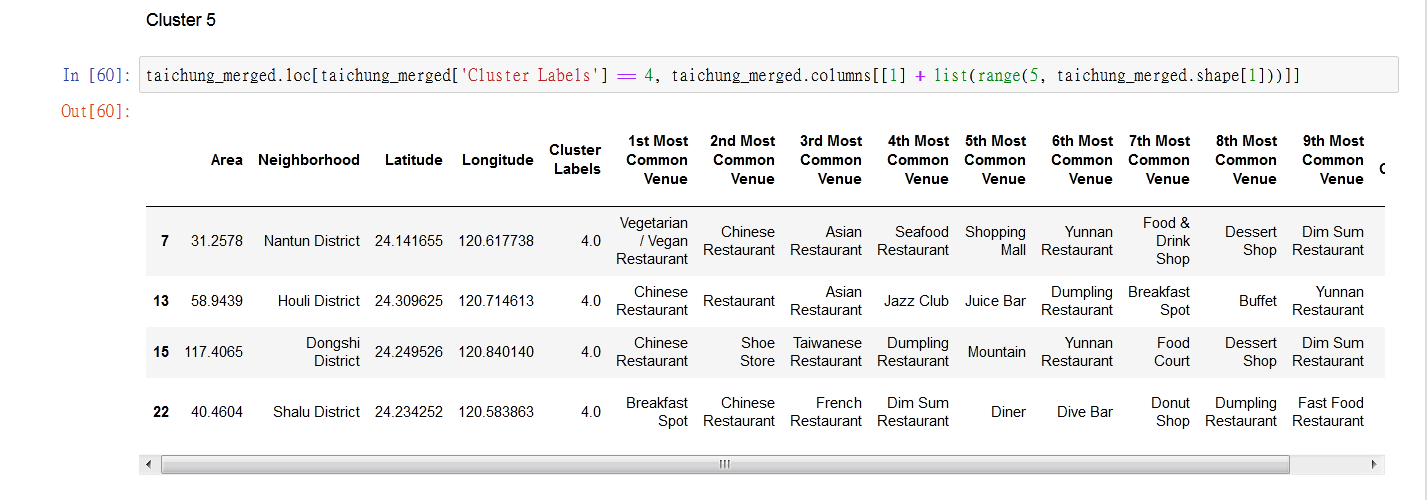


* 1. Examine Clusters









* 1. **Results & Discussion**

1. **Conclusion and Summary**

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