# Introduction

When we decide to visit a country, we may need lots of information pertaining to some target places, i.e. a city, scenery, or national park, etc. In this project, I'd like to develop a system to help tourists visit Taipei City which is the capital and a special municipality of Taiwan. By analyzing the 12 administrative districts in the Taipei City, including Songshan, Xinyi, Daan, Zhongshan, Zhongzheng, Datong, Wanhua, Wenshan, Nangang, Neihu, Shilin, and Beitou, tourists may have initial knowledge of Taipei City which may help them to organize their itinerary. Moreover, when we plan a journey, we may search where to stay with online marketplaces and filter out lots of searches. To help people make a decision more easily, the system will show the regional characteristics with the Foursquare data pertaining to a short list of accommodation choices. More specifically, the project will  provide a function that analyzes the data obtained from Foursquare and assists people to search as well as decide the accommodations when they are traveling, especially, in Taipei.

# Objective

In this project, we will know more about the details of districts in Taipei through machine learning segmentation and clustering. Second, by input a short list of accommodation addresses, we can apply this project to analyze the characteristics of surrounding venues with Foursquare data and K-means and help us to look for accommodations according to our preferences. By means of segmentation and clustering as well as Foursquare data, the following goals are expected to be fulfilled:

1. Analyze the districts in Taipei
2. Recommend the ordered accommodations from the aspect of the specified attributes.

# Data

The district data of Taipei City is scraped from a table on the wikipedia page "https://en.wikipedia.org/wiki/Taipei", which contains the English name, Chinese name, pinyin, population, area, and postal code of each district in Taipei City. After data wrangling, the columns of district data will be English name, Chinese name, population, and area left. For the second objective, a short list of accommodation addresses is prepared in text file format and uploaded to my github for reference "https://github.com/allisonyuplayground/Coursera_Capstone/blob/master/accommodation_list.txt". Each address is stored in a single line of the text file, so that an user can extract a list of addresses by reading the file line by line and storing them into an array. Another aspect of this project is the Foursquare data, such as venue name, venue category, venue latitude, venue longitude, etc., which are mainly used for segmentation and clustering.