Battle of the Neighborhoods

**Introduction**

In this project, I will provide a general guide about where to open a restaurant, a coffee shop, or to setup an office for the new business based on. The Foursquare API is used to explore the neighborhoods in a particular city, and the explore function is used to get the most common venue categories in each neighborhood. After this project, you will get a general idea on determining the location for your business.

**Data**

The data for this project is all the neighborhood in California. There is a total of 55 counties and 481 cities in California. And 500 venues within 1000 meters of each cities are investiaged. Since the total number of venues for all the 55 counties (481 cities) are too big, which exceeds the personal call limit of Foursquare API. I downsize the counties to Log Angeles only, consisting 88 cities, which is famous for its diversity.

**Methodology**

Before the investigation of all the venues in a total of 88 cities in Los Angeles, a plot study focusing on extracting the top 500 venues in the first city in Los Angeles within a radius of 1000 meters is conducted firstly to go through the details and present a clear process of how to get the structured dataset with FourSquare API. The process includes several steps, 1) get the first neighborhood’s name (Agoura Hills), 2) get the url for the Foursqurea API, 3) send the GET request and extract the information step by step, including a) get the overall results, b) get the venues form the key: response, c) flatten JSON and structure it into a pandas dataframe, d) filter columns, extract the columns (or features) of interest, **'venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng'**, e) clear the **venue.categories,** extract the **name** form the categories, f) clear the columns, and remove the unnecessary column **venue.categories**, and 4) visualize the venues in the first city (Agoura Hills) in Los Angeles with Folium map.

After the exploration of all the venues in only one city in Los Angeles, the venues from all 88 cities in Los Angeles are explored. A function (getNearbyVenues) will be created to get the venues of all cities in Los Angeles.

Then the venue category in each city will be analyzed.

Next, all the 88 cities are clustered into five groups based on their characteristics using *k-means* function, the frequency of the venue categories. A *Folium* map clustering all the 88 cities are created with different colors indicating cities belonging to a particular cluster group.

Finally, each of the five clusters are examined separately regarding their venue characteristics.