**Using Data Analytics to recommend new business location**

**January 19, 2019**

**Introduction**

In this project, I will provide a general guide about where you should 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) was created to get the venues of all cities in Los Angeles. A total of 3779 venues (319 unique venue categories, and 2897 unique venues) are identified in Los Angeles. The venue category of Mexican Restaurant is among the top (count =188), followed by the venue category of Fast Food Restaurant (count = 139), Coffee Shop (count = 127), etc. The venues include Starbucks, 7-Eleven, SUBWAY, etc.

Then the venue category in each city was analyzed. Take the first city **Agoura Hills** as the example, the top five most common venues categories are Hotel Bar (0.29), Gym (0.14), American Restaurant (0.14), Hotel (0.14), and Summer Camp (0.14). From the analysis of the venue category in this city, we can explore some of the characteristics of people living there. For example, most Americans are living there compared to the Spanish Restaurant, Mexican Restaurant, Indian Restaurant, or Korean restaurant in other cities.

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.

**Results**

All the 88 cities are clustered into five groups, with 4 cities belonging to the cluster group 1, 36 cities belonging to cluster group 2, one city belonging to cluster group 3, 32 cities belonging to cluster group 4, and five cities belonging to cluster group 5.

**Discussion**

This project provides a rough guide about where to start a business. The data analytics for other key performance indicators (KPIs) are needed before starting a new business.

However, once you determine to start a business, it is recommended to make the start-up lean at first, set the goal on the sand. In this case, you'll not go too far into the bubble you've created, and you won't come out until you hit the wall. At the same time, this lean analytics will help you quantify your innovation, getting you closer and closer to continuous reality check, and build your own business model.

**Conclusion**

Conclusion 1:

From the output above, there are a lot of 88 cities in Los Angeles clustered into five groups based on the characteristics of each group. The numbers of cities for each group are 4, 36, 1, 32, and 5, respectively.

Within 1000 meters for each of the 88 cities, a limit of 300 venues are explores, a total of 3779 venues are recurred. Among the 3779 venues, there are 2897 unique venues belonging to 319 unique venues categories. The venues categories include **Hotel**, **Summer camp**, **Hotel Bar**, **Park**, **Gym**, **American Restaurant**, **Historic Site**, **Monument/Landmark**, **Garden**, and **Palace**, etc.

Conclusion 2:

From the discussion above, the most common venue category for each of the five groups varied a lot, may indicating some of people's living habits in the particular neighborhood. For example, the most common venues categories for group 2 are Fast Food Restaurant and Mexican Restaurant, indicating that people there are living a quite fast-speed life, and Mexicans could occupy a big portion there. Thus, if you want to start a business related to Mexicans, or your business would help improve people's live quality who are living a fast-speed life, you may consider the cities in group 1, 2, or 4.

Moreover, if you want to start a business related to entertainment, such as fishing, diving, pleasure, it could be better to choose your cities belong to group 5.

On the contrary, if you want to start a Chinese Restaurant, it would not be a good choice to start your business in these cities. At least, you need more to do more research focusing on the population of Chinese people there.