Capstone Project - The Battle of the Neighborhoods (Week 2)

A. Introduction / Business Problem

Obesity is a common and serious issue in United States. According to World Atlas, a website that covers demography and sociology topics all over the world, the most obese country in North America is the United States.

Recent years, US citizens have been paying more attention to healthy lifestyle. People are more willing to spend more money in eating healthy food and to live a healthy life. These include spending money in sport activities, signing up for gym membership and hiring of personal trainer.

This provides opportunities for businesses that are offering fitness service, including gym and fitness center. United States is a big country and it has a huge population. United States is also a developed country. Majority of the citizens are able to spend their earning income in areas beside of necessities.

However, United States is such a big country and it is illogical to open new business in every single state, especially for small company and new entrepreneur. Thus, it is very important for us to conduct research before venturing into our Gym Business. With limited resources, we want our target audience to target only the states with the highest opportunities.

Since population is one of the most important factors in determining a successful gym business, we will prepare population data of each city in United States.

Some of the audiences that we are targeting are small business personnel who likes to venture into gym business and medium enterprise which wants to take advantage of gym franchises.

In this project, we will use Foursquare data to obtain some of the most common venues within our targeted cities.

B. Data Preparation

Referring to our problem, factors that will influence the business decission are:
- city population
- per capita income of each city

1. Get the list of United States cities and City Population.

The wikipedia article of "List of United States cities by population" provides the table of United States cities, sorted in order by the city population.

We use BeatifulSoup4 to scrape the table data from the article. We also scrape other data such as population density and land area. Cleaning is required to ensure a neat and useful table. We drop some columns and rename some of the column names. Then, we use pandas library to create the dataframe.

[link to Wikipedia] https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population

2. Get the coordinates of the targeted cities.

Since the population is one of the most important factors in determining a successful gym business, we narrow down our list of cities and only targeted on the three cities that are most populated in United States. Then, we use Geopy library to get the coordinates of these cities. After getting the coordinates, we double check with google search result to ensure the accurateness of the coordinates.

3. Get the Per capita Income Data.

We scrape the data of Per capita Income from the wikipedia article "List of United States counties by per capita income". Per capita income is also one of the important factors in determining a successful gym business. When the income level is high enough, people are willing to spend extra moneny in fitness activities.

[link to Wikipedia]
https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income

4. Foursquare Data

We also use Foursquare Data to get some of the most common venues that can be found in our targeted cities.

	2018rank	City	State[c]	2018estimate	2010Census	2016 land area.1	2016 population density.1
0	1	New York[d]	New York	8398748	8175133	780.9 km2	10,933/km2
1	2	Los Angeles	California	3990456	3792621	1,213.9 km2	3,276/km2
2	3	Chicago	Illinois	2705994	2695598	588.7 km2	4,600/km2

	2018rank	City	State[c]	2018estimate	2010Census	2016 land area.1	2016 population density.1	Per Capita Income	Latitude	Longitude
0	1	New York[d]	New York	8398748	8175133	780.9 km2	10,933/km2	$62,498	40.712728	-74.006015
1	2	Los Angeles	California	3990456	3792621	1,213.9 km2	3,276/km2	$27,749	34.053691	-118.242767
2	3	Chicago	Illinois	2705994	2695598	588.7 km2	4,600/km2	$30,183	41.875562	-87.624421

C. Methodology

After collecting all the required data, let's take about methodology.
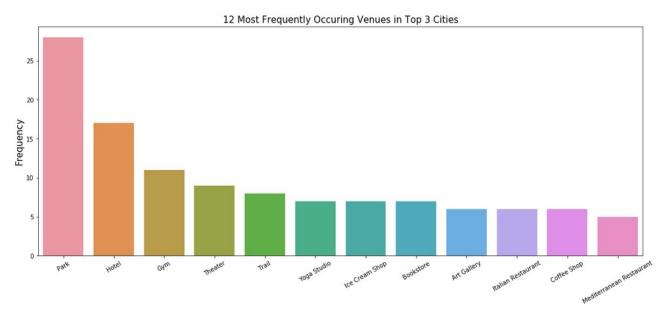
In this project, we will make full use of Foursquare API. We will start by exploring some of the most visited venues in the three cities. We want to check whether gym is one of the most visited venues. Then, we will look for number of gyms in each city. We will also use Hot Encoding to get more information about the venue categories of each city.

D. Analysis

We narrow down our targeted cities to the top three most populated cities. They are New York, Los Angeles and Chicago. Then, we explore the areas around these cities and we focus on gyms. We found that there are a total of 13 gyms in these three cities.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	New York[d]	40.712728	-74.006015	Barry's Bootcamp NoHo	40.728937	-73.992241	Gym
2	New York[d]	40.712728	-74.006015	The Fhitting Room	40.739857	-73.993194	Gym
3	New York[d]	40.712728	-74.006015	Equinox Gramercy	40.740749	-73.985771	Gym
4	New York[d]	40.712728	-74.006015	iLoveKickboxing - Chelsea, NY	40.746340	-73.992900	Gym
5	New York[d]	40.712728	-74.006015	305 Fitness MIDTOWN	40.746790	-73.982885	Gym / Fitness Center
6	Los Angeles	34.053691	-118.242767	LA Boulders	34.038892	-118.234508	Climbing Gym
7	Los Angeles	34.053691	-118.242767	Equinox Hollywood	34.099440	-118.326627	Gym
8	Los Angeles	34.053691	-118.242767	Cliffs of Id	34.032957	-118.370943	Climbing Gym
9	Los Angeles	34.053691	-118.242767	Los Angeles School of Gymnastics	34.022020	-118.379456	Gymnastics Gym
10	Chicago	41.875562	-87.624421	Studio Three	41.893999	-87.631373	Gym / Fitness Center
11	Chicago	41.875562	-87.624421	East Bank Club	41.889694	-87.638922	Gym

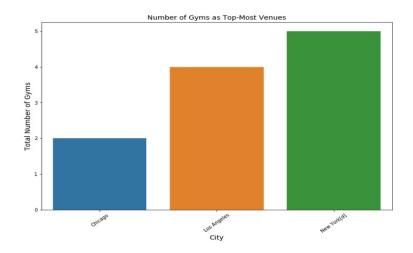We also found that Parks top the bar charts of most common venues in the three cities, followed by Hotel, Gym and Theater.

12 Most Frequently Occuring Venues in Top 3 Cities

Next, we find the top five venues of each city.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Chicago	Hotel	Park	Theater	Italian Restaurant	Yoga Studio
2	Los Angeles	Hotel	Trail	Theme Park	Park	Coffee Shop
3	New York[d]	Park	Bookstore	Gym	Cycle Studio	Art Gallery

We also find out that the number of gyms as top most venues in each city.

Number of Gyms as Top-Most Venues

E. Result & Discussion

After analysis, let's take a look at our results. We have used data from Wikipedia and Foursquare API, and we have found that:

1. Park is the most common venues in the three cities.
2. Hotel is dominating Chicago and Los Angeles as the most common venue.
3. Park is the top common venue in New York.
4. New York has the most number of gyms, while Chicago has the least number of gyms.
5. New York is the most populated city and Chicago is the least populated citiy among the three cities.
6. New York has the highest per capita income.

Based on the analysis, Chicago has a less competitive environment compared to the other two cities, as Chicago has the least number of gyms. Besides, although Chicago is the least populated city among the three cities, it has higher per capita income and higher population density than Los Angeles. This indicates that if the new gym business can be set up in a strategic location in Chicago, it can exploit the advantage of its population density. Chicago citizens also have higher per capita income than Los Angeles citizens. They could be more willingly in spending extra earning in fitness activities.

Another finding is although gym does not top the chart of most common venues, we can find that park is dominating the chart. It means citizens are spending their time in parks. It could be a walk, a jog or any recreational activity. This indicates a trend where citizens are actually paying attention to healthy lifestyle by doing exercises in parks. Health- conscious citizens provide opportunities to gym business.

F. Conclusion

This project serves its purpose to aid potential entrepreneur to look for a strategic location and goal for his gym business. By finding out population data, per capita income and most common venues of our targeted cities, we have identified the most suitable city for gym business.

However, entrepreneur will make the final decision. Although Chicago is the least competitive market and it seems like the most suitable city for setting up new gym business, entrepreneur can always go for New York if he or she has adequate resources. New York might be the most competitive market but if one can manage to stand out among competitors, New York as the most populated city and has the highest per capita income can definitely provide unlimited opportunities.