

# Where in China is a good place to open a sport shop?

## Introduction

This PDF file is a report of the Applied Capstone Project – IBM via Coursera, where a key question of **“If I want to open a sports shop in China, where would be a good place to do it?”** is answered.

## Background

The Covid-19 pandemic has shaken the entire world since Dec 2019. 7 months later, some parts of the globe such as Europe and Asian has survived from the first shock. Many of us, including myself, has learned an important message from this tragical worldwide event, and that is, health related issue is the top priority, and it does come before anything in life. In the spirit of this, the author believes many people will value their health more than ever in a post-Covid-19 society. Take the UK Prime minister Boris Johnson as an example, he himself has a BMI of a slight overweight person and has battled Covid-19 with the help of NHS. Surprisingly, he started to exercise via running in the national parks after being discharged from the ICU. Hence, logically speaking, the demand of local health related venues like gym and fitness shall rise accordingly.

From a business perspective, an increase of the local fitness/gym venues is usually accompanied by an increase in the sport shops being opened around. As China has already recovered from Covid-19 since April 2020, in addition, with the society being re-opened, their 2<sup>nd</sup> quarter national GDP has reversed from being negative to positive 3.2%.<sup>[1]</sup> Therefore, the author decided to define the business question of **“In the spirit of believing there will be an increase demand of health related venues such as gym and fitness centre, I wish to open a sport shop in China, where would be a good location to do so?”**

The prior stakeholders who might be interested in this analysis would be parties or individuals that wish to open gym/fitness related accessories shops such as Nike, Adidas and Decathlon. The analysis within this report would also be quite interesting for those who wish to do a concrete regional research of sports leisure shops and its relevant by-products in some of the major cities in China.

## Data

To figure out where in China is a good place to open a sport shop, the author created a thinking process to demonstrate the steps and the required data that are needed to solve this question. Please see the figure below.

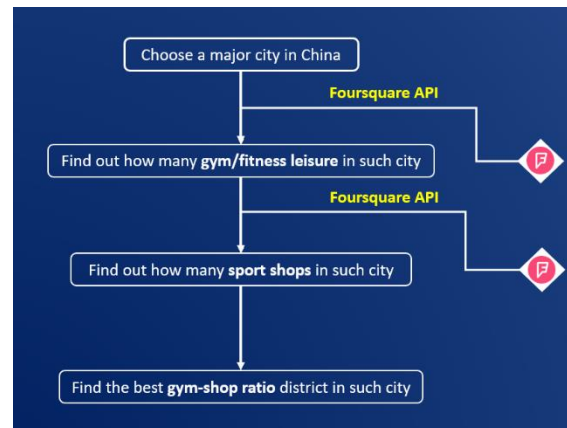


Figure 1: thinking process

As shown, the first step is to choose a major city in China. As China has 22 provinces with many cities, the author decided to analyse 4 provinces' capital cities to get a good comparison. The chosen cities are shown in the following map below. \*[The choice of these cities is based on a random selection of the “first line” and “super first line” cities in China. Some websites also called it Tier 1 and Tier 2 cities, see reference 2 for more details of the classification of the cities in China] <sup>[2]</sup>



Figure 2: major cities maps

The second step as shown is to find out how many gym/fitness venues are there in the chosen cities. To do so, the author uses an independent location data platform called the Foursquare.com. Via the Foursquare API, information such as location in the form of latitude, longitude coordinates and address; tips; name of each fitness/gym venues are returned.

As for the third step, similarly, within each of the chosen cities, the quantity of sports-accessories related shops; their relevant information such as its location and distribution are also acquired via the Foursquare API. The last step of this analysis is quite straightforward, under the assumption of “a good location to open a sport shop is when there is an abundant number of gym/fitness venues locally with insufficient or relatively low quantity of sports shops around.”, a defined index of the gym/fitness to shops ratio [gym/fitness : shops] are calculated.

Therefore, the main datasets that are required to analyse the ideal location to open such shop are:

- 1) Major city location data [lat, lng]
- 2) Gym/fitness venues info in each city
- 3) Sports shop info in each city

The detailed datasets and how they are used are illustrated in the methodology session of this report.

## Methodology - (Python via Jupyter)

### 1. Import libraries

As the major cities [Beijing, Shanghai, Chengdu, Chongqing] are chosen by the author, the first step in this analysis is to import the libraries that are needed in Python via Jupyter notebook to perform the analysis. The libraries that are going to be used are:

- 1) Pandas - data analysis in tabular form
- 2) Numpy - vectorized data for processing
- 3) Json\_normalize - deal with .json file
- 4) Folium - map display
- 5) Geopy - location tools

\*Remark, the programming software used in this analysis is Jupyter – Python 3.0, see github link for a detailed code display.

```
import pandas as pd
import numpy as np
from pandas.io.json import json_normalize
import folium
from geopy.geocoders import Nominatim
print('Libraries imported')
```

Libraries imported

Figure 3: import libraries

### 2. Generate location info for cities

To display a map of the major cities that are chosen, it is needed to generate the latitude and longitude coordinates for each city. To do so, we use the attributes from geopy (library to generate location data) [geo.geocoders.Nominatim] as shown below:

```
geolocator.Beijing = Nominatim(user_agent = 'TayTay')
city_1 = 'Beijing'
country = 'China'
location_Beijing = geolocator.Beijing.geocode(city_1 + ',' + country)
print(location_Beijing.address)
print(location_Beijing.latitude, location_Beijing.longitude)
```

北京市, 东城区, 北京市, 100010, China 中国  
39.906217 116.3912757

```
geolocator.Shanghai = Nominatim(user_agent = 'TayTay')
city_2 = 'Shanghai'
country = 'China'
location_Shanghai = geolocator.Shanghai.geocode(city_2 + ',' + country)
print(location_Shanghai.address)
print(location_Shanghai.latitude, location_Shanghai.longitude)
```

上海市, 黄浦区, 上海市, 200010, China 中国  
31.2252985 121.4890497

```
geolocator.Chengdu = Nominatim(user_agent = 'TayTay')
city_3 = 'Chengdu'
country = 'China'
location_Chengdu = geolocator.Chengdu.geocode(city_3 + ',' + country)
print(location_Chengdu.address)
print(location_Chengdu.latitude, location_Chengdu.longitude)
```

成都市, 四川省, China 中国  
30.6624205 104.0633219

```
geolocator.Chongqing = Nominatim(user_agent = 'TayTay')
city_4 = 'Chongqing'
country = 'China'
location_Chongqing = geolocator.Chongqing.geocode(city_4 + ',' + country)
print(location_Chongqing.address)
print(location_Chongqing.latitude, location_Chongqing.longitude)
```

重庆市, 渝中区 (Yuzhong), 重庆市中心, 重庆市, 400014, China 中国  
29.5585712 106.5492822

Figure 4: location coordinates

### 3. Map generation for demonstration

After the location information [latitude, longitude] for each city being generated via geopy, we use folium (a popular python library for map display) to display maps for each city serve for demonstration purpose. Here we take Beijing as an example.

```
#round coordinates into 2 decimal places
latitude_0 = location_Beijing.latitude
longitude_0 = location_Beijing.longitude
longitude_0_0 = round(longitude_0,2)
latitude_0_0 = round(latitude_0,2)

latitude_1 = location_Shanghai.latitude
longitude_1 = location_Shanghai.longitude
longitude_1_1 = round(longitude_1,2)
latitude_1_1 = round(latitude_1,2)

latitude_2 = location_Chengdu.latitude
longitude_2 = location_Chengdu.longitude
longitude_2_2 = round(longitude_2,2)
latitude_2_2 = round(latitude_2,2)

latitude_3 = location_Chongqing.latitude
longitude_3 = location_Chongqing.longitude
longitude_3_3 = round(longitude_3,2)
latitude_3_3 = round(latitude_3,2)
```

Figure 5: D.C rounding

Note that in the github code scripts, we rounded the decimal places of the location coordinates into 2 decimal places. Unfortunately, folium does not display location in China in English, if you wish to confirm the location map of Beijing, please try use the latitude, longitude coordinates in google map.



Figure 6: map of Beijing, China

## 4. Local venue exploration

Now we enter the fun parts!

To explore the local gym/fitness venues, Foursquare API is used by the author to get location data of each venue. After registering a free/trail account for developers at Foursquare.com, a client id; client secret is created. Using these “keys”, a connection to foursquare search can be made via python. The foursquare client id and secret’s form look like below, and the version usually can be defined by the date of each search. [ the last 4 digits of the author’s ID is blurred for security purposes]:

```
CLIENT_ID =
'P1X3FJMKPO4DWCK1FDOBLB51PF4QPP4C1OQIQ
ENXUREV0000'

CLIENT_SECRET =
'VD1ZTKQOQOKTIONJ5TGC0VXF0DK25YYOBOINQ13
JONET00000'

VERSION = '20200727'
```

Seeing the code in figure 7, we create a selection of json files that contains the results from the search via Foursquare API, the information that were needed for such search were the Foursquare API details, the version usually indicated the date of search, the cities that were chosen, and the category ID of venues search. To see a full list of the category ID of various venues please press Ctrl and click on the reference link here. [\[3\]](#)

```
import requests

LIMIT = 1000
cities = ['Beijing,Beijing', 'Shanghai, Shanghai', 'Chengdu, Sichuan', \
'Chongqing, Chongqing']
results_gym = {}
for city in cities:
    url = "https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&version={}&city={}&limit={}&category_id={}"
    results_gym[city] = requests.get(url).json()
```

Figure 7: Foursquare search

The json file of each city’s results usually looks like a big dictionary of information, see below:

```
In [60]: results_gym[cities[0]]

Out[60]: {'meta': {'code': 200, 'requestId': '5f1ebfbcb37754cc7ab5b3'},
'response': {'geocode': {'cc': 'CN',
'center': {'lat': 39.9075, 'lng': 116.39723},
'displayString': 'Beijing, Beijing Shi, China',
'geometry': {'bounds': {'ne': {'lat': 41.05875499151485,
'lng': 117.49936797940963},
'sw': {'lat': 39.44236100189344, 'lng': 115.41642409091048}}},
'longId': '72057594039744606',
'slug': 'beijing',
'what': '',
'where': 'beijing beijing'},
'groups': [{'items': [{'reasons': {'count': 0,
'items': [{'reasonName': 'globalInteractionReason',
'summary': 'This spot is popular',
'type': 'general'}]},
'referenceId': 'e-0-4da2c281540e1cd759382de-0',
'venue': {'categories': [{'icon': 'prefix': 'https://ss3.4sqi.net/im
g/categories_v2/building/gym',
'suffix': '.png']}]}]}]}
```

Figure 8: gym/fitness results

The information in the above json. file looks rather unorganised and difficult to understand for non-developers, but that is not an issue. We can extract the information that we are interested in by simply transforming it into a pandas data frame, i.e. a tabular form of rows and columns. In this case, we are interested in the location data of each fitness venue in the form of address, geo coordinates and the name. The geo coordinates, i.e. latitude and longitude are useful when we wish to add markers on the maps to display where are the venues.

```
df_venues = {}
for city in cities:
    venues = json_normalize(results_gym[city]['response']['groups'][0]['items'])
    df_venues[city] = venues[['venue.name', 'venue.location.address', \
'venue.location.lat', 'venue.location.lng']]
    df_venues[city].columns = ['Name', 'Address', 'Lat', 'Lng']

df_venues[cities[0]].head(10)
```

	Name	Address	Lat	Lng
0	Kerry Sports Center (高里健身俱乐部)	1 Guanghua Rd	39.912542	116.453264
1	Pacific Century Club	2A GongtiBei Rd.	39.930683	116.452938
2	B Active Sanlitun SOHO	NaN	39.930395	116.447208
3	D-Space 东田运动	Futong West Street	39.997054	116.474809
4	Nirvana Fitness & Spa	2 Workers Stadium N Rd	39.932234	116.453520
5	宝力豪健身 PowerHouse Gym	东直门外大街40号 东方银座写字楼C层	39.939809	116.429474
6	WILL'S 威尔士健身	B1-26, U-TOWN Mall	39.919961	116.433128
7	Keepland Fortune Plaza 财富购物中心店	财富购物中心 2F	39.916225	116.453762
8	Amrita Fitness (瑞士酒店健身中心)	No. 2 Chao Yang Men Bei Dai Jie	39.932196	116.428953
9	Joy sunny Fitness	天之骄子	39.893725	116.458942

Figure 9: transfer json info into pandas data frame

As shown above in figure 9, we created a table and transferred the information from our json file for each city’s results in term of local gym/fitness venue, in such table, it is much clearer to view the venues details. However, although the location data with the addresses for each venue is convenient to see, without a map to visualize where the gyms are in each city, the audience still does not have a good picture of the density and distribution of the venues. Hence, folium is used here to display each gym in the chosen cities.

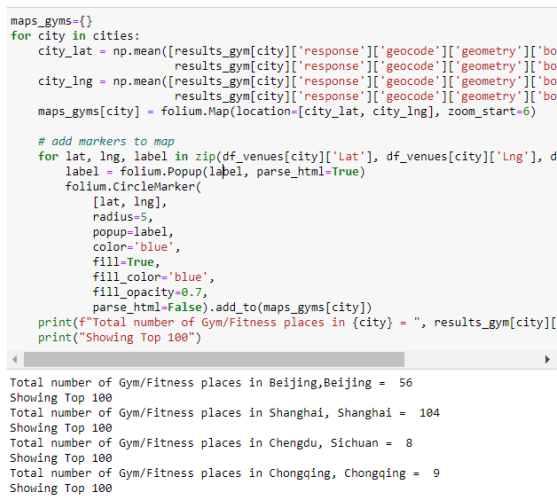


Figure 10: Map creation

As shown in the code above, for each gym/fitness venues in each city, a map with marker indicates the location, lng, each venue is created. In addition, the total number of gym/fitness venue for each city is calculated and displayed. See the following figure for all gym/fitness venues searched via Foursquare in Beijing as an example. The map display for the rest of the cities are in the code scripts on github.

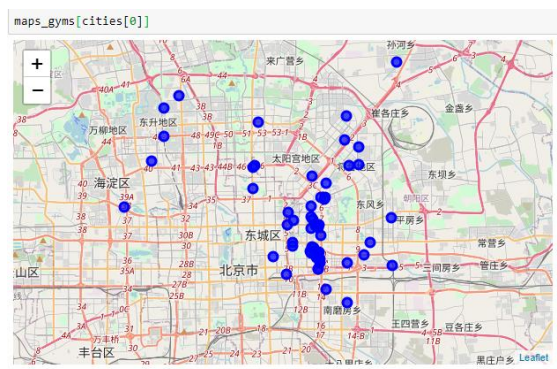


Figure 11: Local gym/fitness in Beijing

## 5. local sports shop exploration

Now, similarly, we want to explore how many sports shops are there in each chosen city. Using the same method in the above session with only a different category ID for the sports shops, a json file for each city's results of local sports shop is created.

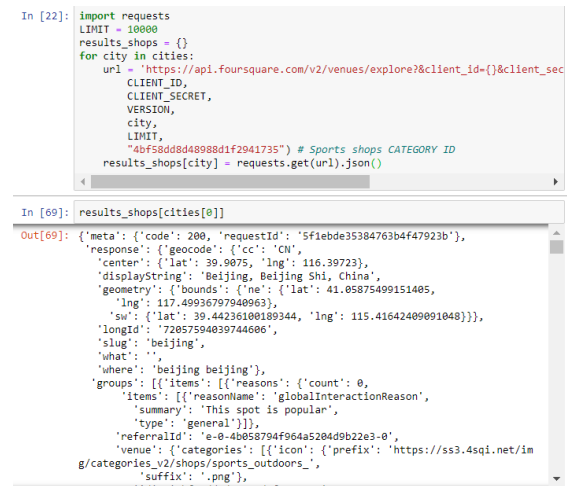


Figure 12: Local sports shops search in Beijing

Same as the previous json files of the results of local gym/fitness venues, the original data of each sports shops would be easier to comprehend in a pandas data frame. In figure 13, the name, category and the geo locational coordinates of each sports shop in Beijing are listed.

	name	categories	lat	lng
0	Decathlon (迪卡侬)	Sporting Goods Shop	39.884797	116.482460
1	Decathlon (迪卡侬)	Sporting Goods Shop	39.977370	116.454598
2	adidas	Sporting Goods Shop	39.932521	116.448381
3	Nike	Sporting Goods Shop	39.933338	116.448301
4	Decathlon 迪卡侬	Sporting Goods Shop	40.014956	116.457887
5	Decathlon (迪卡侬)	Sporting Goods Shop	40.058611	116.409275
6	lululemon	Sporting Goods Shop	39.937177	116.448282
7	Decathlon (迪卡侬)	Sporting Goods Shop	39.952006	116.350091
8	Nike Basketball & Jordan Wukesong	Sporting Goods Shop	39.908175	116.271507
9	李宁 Li-Ning	Sporting Goods Shop	39.911185	116.405035
10	Decathlon (迪卡侬)	Sporting Goods Shop	40.045487	116.330995

Figure 13: Local sports shops in Beijing

As shown in the above table, typically, shops like Nike, Adidas, Decathlons are quite abundant locally. This information would be quite interesting for a competitor analysis for stakeholders such as newly branched sports leisure shops, however, that is not the prior interest of this report.

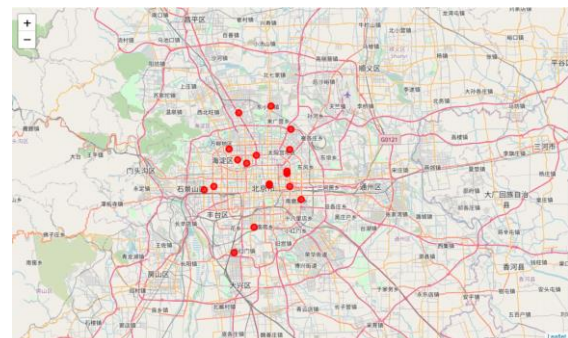


Figure 14: Local sports shops in Beijing

The total number of sports shops in each city is also calculated and printed via the print function. Again,



a map display example of Beijing is shown in the figure 14. Note that the local venues in the maps are displayed in blue circular solid marker, and the local sports shops are displayed in red circular marker.

## Results & Discussion

As shown in the table below, the number of gyms in Beijing and Shanghai are far more than Chengdu and Chongqing. Hence, we first narrow down the cities of choice to Beijing and Shanghai.

Recall from previous session, after we explored the local gym and sports shop venues, a straightforward calculation of the index of the gym to shops ratio for Beijing and Shanghai is calculated.

City	No.of Gyms	No. of Shops
Beijing	56	21
Shanghai	104	43
Chengdu	8	16
Chongqing	9	9

[\*Remark, as Foursquare.com is an English data platform, the gym/fitness/shops under name of Chinese only are not displayed here.]

```
no_of_gyms_Beijing = results_gym[cities[0]]['response']['totalResults']
no_of_shops_Beijing = results_shops[cities[0]]['response']['totalResults']
gym_shops_ratio_Beijing = no_of_gyms_Beijing/no_of_shops_Beijing
print(gym_shops_ratio_Beijing)

2.3333333333333335

no_of_gyms_Shanghai = results_gym[cities[1]]['response']['totalResults']
no_of_shops_Shanghai = results_shops[cities[1]]['response']['totalResults']
gym_shops_ratio_Shanghai = no_of_gyms_Shanghai/no_of_shops_Shanghai
print(gym_shops_ratio_Shanghai)

2.4186046511627906
```

Figure 15: index ratio for Beijing and Shanghai

As shown in the below figure, the index ratio for Beijing and Shanghai is quite comparable, however, the number of gym and sports shops in Shanghai almost doubled its amount compare to Beijing. Judging by the number of local gym and sports shops, an indication of “people in Shanghai seems to have a higher demand in sports related leisure” is drawn. Hence, the author decided to narrow down the city choice to Shanghai.

Then, let us take a closer look at the distribution of the local gyms in Shanghai. Seeing from Figure 16, the local gyms in Shanghai seems to be quite dense in the area of few districts.

1. ChangNing district – circled in black;
2. JingAn district in Yellow;
3. HuangPu district in Orange

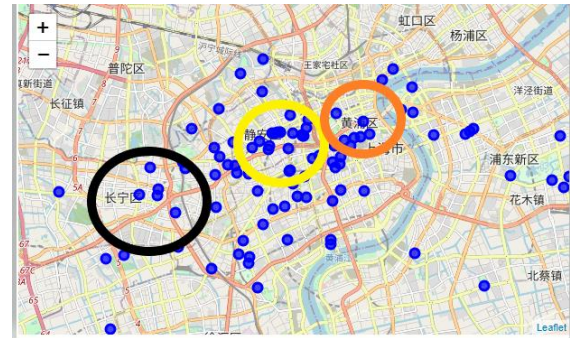


Figure 16: District in Shanghai with dense gyms distribution

It came across to the author that after zoomed in to visually check for each district, there is a very special distribution of the gyms on one road of JingAn district, where 5 gyms are opened in one single road, and it is a unique scenario and no other areas in Shanghai has the equivalent. See figure below.



Figure 17: Unique distribution in JingAn district

The next step is to check the local shops distribution, for convenience, a map with markers indicate both gym and shops in Shanghai is created. Where blue marker indicates the local gym, and red marker pins the local sport shops. For the target area shown in figure 17, there are only 3 sports shops and 14 gyms within an area of 1000m radius (orange circle). A black circle indicates an area of 800m radius is also drawn to show that there are only 3 sports shops and 11 gyms.

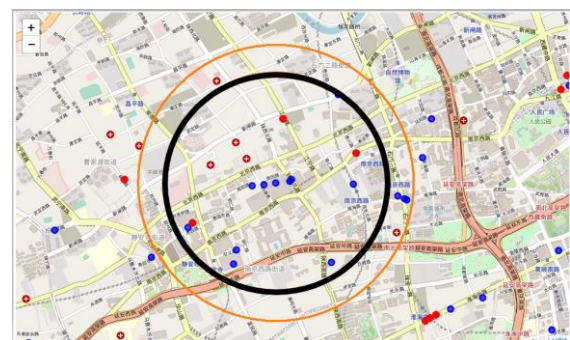


Figure 18: Target area

Therefore, the author has chosen the area of the orange circle within JingAn district in Shanghai to be a good location to open a sports shop.

Although there are other areas in Shanghai where there is an abundant number of gyms with no sports shops, the author still chose the area mentioned above as people tend to shop in areas where there is a various selection of shops. And this is fairly easy to understand, when people going out for shopping, they either tend to go to the local high street where there are many shops of different categories or big shopping malls, it is rare that people shop in an area with no shops.

Another point of worth discussing is that, as seen in the tables for local gyms and shops, the number is quite low. This is because that Foursquare API only returns results which has English translation, for those shops and gyms without English information, the search engine will not return it as a result. If google API or trip advisor is used, the number of shops and gyms would be higher. However, for the purpose of demonstration, Foursquare API is quite sufficient.

In conclusion, after comparing the local sports shops and the gym/fitness venues in 4 major cities in China, under the assumption of "A good location to open a sport shop in China is where there is an abundant number of gym/fitness venues with a relatively low competitors", the author chose the area in JingAn district in Shanghai to be a good location to open a sport shop.

The relevant code script is available on github, the references used in this analysis is available in each link mentioned, to view it, just simply press ctrl and left click the reference index. Thanks for reviewing!