Capstone Project – Where To Start My Business

Introduction

To young freshmen who are new to workplace and society, instead of working in a firm and becoming an employee, many of them may choose to start their own businesses. Not only for young generations, after working for several firms or being an employee of a particular job position, there are many people who want to be entrepreneur, in order to take control on their career.

Business environment varies among different countries, to decide which country is suitable for this project, I simply take comparison of GDP as my beginning point. Data used will be explained in the next section. After comparison, it is decided to use my home country, China, as my target country and finish my plan.

By doing further analysis, in this project, I will show my result of where is the most suitable place to start a retail business (let's begin with a retail shop first).

Data

GDP of different countries

To decide which country to start with, I use table from Wikipedia to compare the GDP among different countries. As GDP is the most representative indicator showing the economic power of a place, its data is reliable and easy to collect.

Per-Capita GDP of Chinese provinces

As China is the target country in my project, to define the suitable venue more precisely, it is needed to narrow the criteria. Same as above, but this time, per-capita GDP may be more suitable to reflect citizens' purchasing power in a city. Therefore, I have chosen per-capita GDP of Chinese provinces as my required data. After the ranking, it is easier to choose my target province.

Per-Capita GDP of Prefecture-level city in Guang Dong Province

After analyzing, Guang Dong province has been chosen, same as above, per-capita GDP is an useful indicator to narrow the criteria and helps me to do a more precise decision.

Foursquare location data

Now, the final decision is to start retail business in Shen Zhen. But how to choose the right location in this city? Foursquare provides information on Shen Zhen, in this case, since I want to start a retail business, the most related and suitable type of data is 'shopping mall'. Size of shopping mall sometimes represent the popularity of a district, Foursquare helps to present the information of shopping mall in Shen Zhen, so that I can make a conclusion

Methodology

Scraping data from Wikipedia using API

- Beautiful Soup

Beautiful soup is a package to extract data from HTML. In my project, it is used to extract data from Wikipedia and create data frame in further steps.

```
# Getting GDP Data among different countries
response=requests.get('https://www.worldometers.info/gdp/gdp-by-country/').text
soup = BeautifulSoup(response, "lxml")
print(soup.prettify())
```

After codes in the required webpage is presented, it is very convenient to choose 'table' using 'find.all' command, and to create data frame using targeted data.

Presenting data in data frame format

Data Frame is a table showing data in different categories, by analyzing data and putting them into a frame, it is clear to see the result.
Also, I can set my own data frame by creating dictionary and list.

For example:

```
gdpdf = pd.read_html(str(gdptable))[0]
gdpdf.head(5)
```

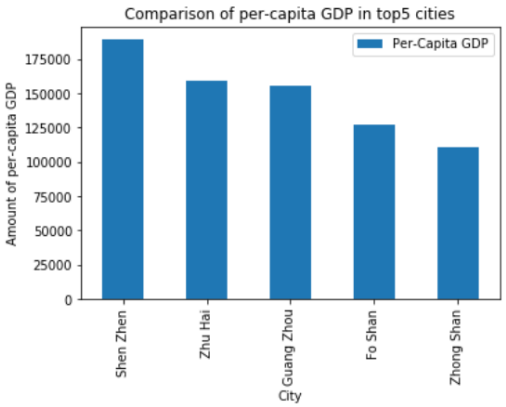
]:

	#	Country	GDP (nominal, 2017)	GDP (abbrev.)	GDP growth	Population (2017)	GDP per capita	Share of World GDP
0	1	United States	$19,485,394,000,000	$19.485 trillion	2.27%	325084756	$59,939	24.08%
1	2	China	$12,237,700,479,375	$12.238 trillion	6.90%	1421021791	$8,612	15.12%
2	3	Japan	$4,872,415,104,315	$4.872 trillion	1.71%	127502725	$38,214	6.02%
3	4	Germany	$3,693,204,332,230	$3.693 trillion	2.22%	82658409	$44,680	4.56%
4	5	India	$2,650,725,335,364	$2.651 trillion	6.68%	1338676785	$1,980	3.28%

Data Visualization-Bar Chart

Sometimes, visualizing data is necessary because it does help audiences to evaluate information in a more convenience way. In order to compare the per-capita GDP in the top 5 cities in Guang Dong Province, I use bar chart to visualize my data, in instead of just selecting top 5 rows from data frame.

```
top5city=top5city.astype(float)
top5city.plot(kind='bar')
plt.xlabel('City')
plt.ylabel('Amount of per-capita GDP')
plt.title('Comparison of per-capita GDP in top5 cities')
plt.show()
```



Leveraging Foursqaure data

Foursquare is a website to get information on a particular place, it provides details on different spots. In my project, I am choosing a suitable place to start my retail business, for example, opening my own store. Therefore, all I have to focus is the shopping malls in my targeted city Shen Zhen. To leverage data in Foursquare, I entered the basic data of Shen Zhen (i.e. latitude and longitude)

```
address = 'Shen Zhen, China'

geolocator = Nominatim(user_agent="foursquare_agent")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print(latitude, longitude)
```

```
22.555454 114.0543297
```

```
search_query = 'Shopping'
url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&v={}&query={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, latitude, long
url
```

```
]: 'https://api.foursquare.com/v2/venues/search?client_id=RTACJTCMCJBZ2OHRMPIRZDKKW4MYSMCO1XCV233YCCUF2JXD&client_secret=4HZST1A0TWYN323MHHDTMNEYUX1JJVCFM5XZJIJ
   P3OH2D4EK&ll=22.555454,114.0543297&v=20180604&query=Shopping&limit=30'
```

However, sometimes the website cannot give me a comprehensive analysis of a place since we still have to rely on others' recommendations. In this case, I choose to read tips provided by another user. Although there is one tip only, I believe that it does help in my conclusion.

	text	agreeCount	disagreeCount	id	user.firstName	user.lastName	user.gender	user.id
0	Nothing much here except bars, if you want to shop go coco park across the road.	2	0	5544d58c498e5021a56650d2	William	K	NaN	2938446

Results and conclusion

In this project, my aim is to choose a place to start my retail business. The criteria is very large, so I begin with choosing country, after that I can detail my plan by narrowing choices.
By comparing GDP among different countries, although USA has the highest GDP, the GDP growth of China seems to be more attractive, it means this country may bring higher profitability to my business.
In the majority of my analysis, I rely on economic indicator, such as GDP and per-capita GDP, to finish my analysis. They are public, transparent to every user and being reliable at the mean time. It is difficult to get detailed consumption level in different places, therefore, the indicators I used might be the most appropriate option.
Moving on to Foursquare part, I choose to scrap only the 'shopping mall' related data in Shen Zhen. It is unfortunate that there is only one tip provided for my target. Therefore, I have to admit that the choice I made may not be the best one in reality, but based on the data I get, it is not bad.

To conclude, collecting and analyzing data from a wider aspect help me a lot in narrowing criteria, and the final answer to my question, which is where to start my retail business, is Shen Zhen.