Capstone Project

The Battle of Neighborhoods (Week 2)

Created by Tien Hoang

Powered by Wikipedia, Google and Foursquare

This project is about findding the good location to open the coffee or milk tea shop in the middle of the tourist city, Budapest. This project can be a good base to be upgrade latter, for example, finding locations to build a coffee chain.

1. Introduction: Business Problem

I am a student who currently study in Hungary. And after studying, I want to start-up my own business. As a milk tea lover, I do love to run my own milk tea shop located in **Budapest**. And with this project, I will try to find the optimal location for the **coffee shop** in this tourist city.

The target is to find a best location which **attract a lot of tourist** and **not already crowded with coffee shops**. We also prefer to locate the shop near the city center as possible in order to serve the Asian tourist, for example Chinese, Vietnamese because of this is their favorite drink. Then latter on, I will try to broaden this milk tea chain around Budapest.

Thanks for the helps of data science, I will find the good place for our milk tea shop.

2. Data

To find out the best location in Budapest, I might need to access to some data. My target is too make this project can be reuse later, so I will use easy access data like **Wikipedia** and **Foursquare location search**

Based on the problem, the direct factors can be:

- Number of existing coffee shop in the neighborhood (since there are just a few milk tea store in Budapest, the direct competitors in this case are coffee shops)
- Distance of neighborhood from city center and attractive tourist places.

The data should be use are:

- **Wikipedia** search about Budapest tourist places: 'https://en.m.wikipedia.org/wiki/List_of_tourist_attractions_in_Budapest'
- The information of nearby coffee shops from **Foursquare API**
- In order to attract the location coordinate, **Google Maps API geocoding** will be used in this project

2.1. Data collection

Installing and importing packages and library

```python
from bs4 import BeautifulSoup
import pandas as pd
import requests
```

```python
url = 'https://en.m.wikipedia.org/wiki/List_of_tourist_attractions_in_Budapest'
source = requests.get(url).text
data = BeautifulSoup(source,'lxml')
```

```python
# define the dataframe columns
column_names = ['Name']

# instantiate the dataframe
budapest = pd.DataFrame(columns=column_names)
budapest
```

Extract name of tourist location from the wikipedia page

```python
#find tourist location in Budapest
name=[]
content = data.findAll("tr", {"bgcolor" : "#DDEEFF"})
for tr in content:
    name= tr.find('b')
    for b in name:
        location = name.find('a')
        if location != None:
            budapest = budapest.append({'Name':location.get_text()},ignore_index=True)
df=budapest
budapest[0:5]
```

	Name
0	Parliament
1	St. Stephen's Basilica
2	Hungarian National Museum
3	National Theatre
4	Buda Castle

Finding the latitude and longitude for these location with Google API

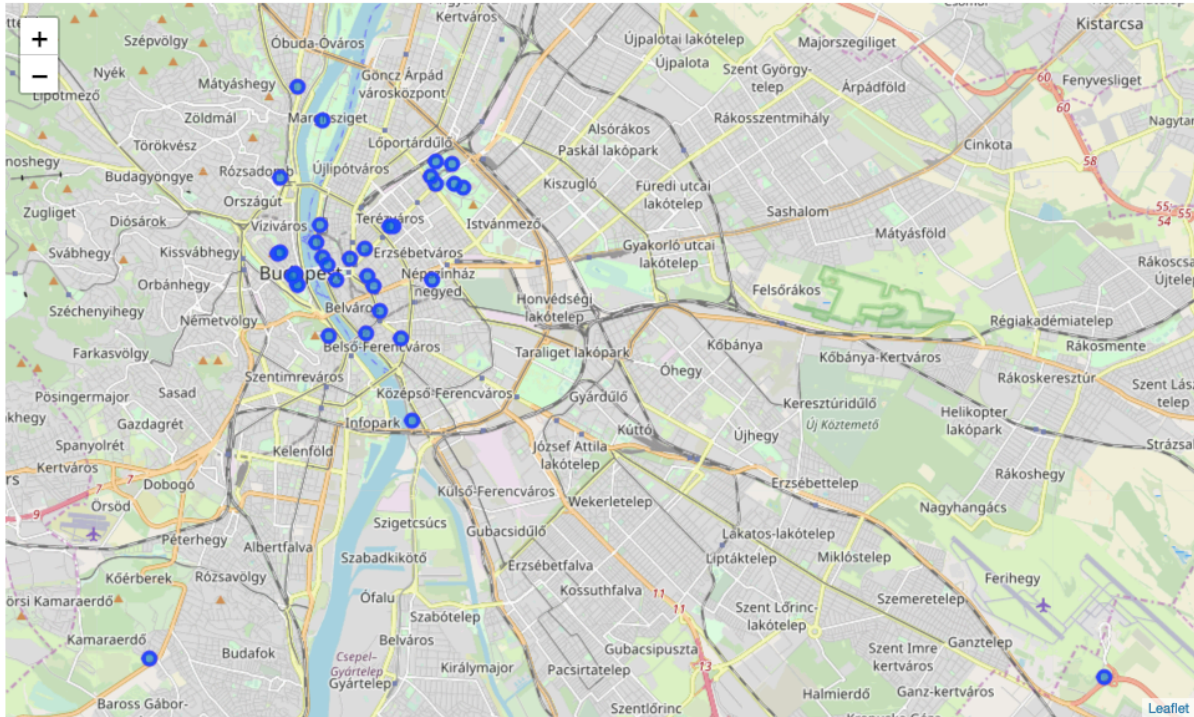|	Name	Latitude	Longitude
0	Parliament	47.507121	19.045669
1	St. Stephen's Basilica	47.500890	19.053989
2	Hungarian National Museum	47.491198	19.062476
3	National Theatre	47.470798	19.071066
4	Buda Castle	47.496205	19.039567

3. Methodology¶

The first requirement of this project is solving the problem by applying K-means clustering to perform city segmentation. With the part of the city being clustered into clusters, the task will become finding the good location which near the attractive place and have less coffee shops.

The Second and Third requirement can be solved by taking venue information from Foursquare API, the location can be traced back with the help of Google Geo API.

```
#Location of Budapest
address = 'Budapest'

geolocator = Nominatim()
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Budapest are {}, {}.'.format(latitude, longitude))
```

The geograpical coordinate of Budapest are 47.4983815, 19.0404707.

Then with the helps of Foursquare API, neighborhood of these point was researched with the radius of 500 m and the limit of 100. The following table is an example for famous places near Parlament, Budapest.

	name	categories	lat	lng
0	Parlament	Capitol Building	47.507041	19.045658
1	Kossuth Lajos tér	Plaza	47.506896	19.047107
2	1956. In Memorian Kossuth tér	History Museum	47.506370	19.046642
3	Bortársaság	Wine Shop	47.505649	19.048543
4	Madal Cafe - Espresso & Brew Bar	Coffee Shop	47.507349	19.049045
5	Culinaris	Gourmet Shop	47.509927	19.046944
6	Magnolia Day Spa	Spa	47.504171	19.045808
7	Cipők a Duna-parton	Outdoor Sculpture	47.503843	19.044917
8	Iguana Bar & Grill	Mexican Restaurant	47.504563	19.048831
9	Da Mario	Italian Restaurant	47.505221	19.049208

The Budapest venues are listed here:

1. Parliament
2. St. Stephen's Basilica
3. Hungarian National Museum

4. National Theatre
5. Buda Castle
6. Hungarian National Gallery
7. Matthias Church
8. Fisherman's Bastion
9. Castle Hill Funicular
10. Sándor Palace
11. Aquincum Military Amphitheatre
12. Hungarian Academy of Sciences
13. Shoes on the Danube Bank memorial
14. Gresham Palace
15. Tomb of Gül Baba
16. Andrássy Avenue
17. Hungarian State Opera House
18. House of Terror Museum
19. City Park
20. Heroes' Square
21. Vajdahunyad Castle
22. Museum of Fine Arts
23. Zoo
24. Széchenyi Medicinal Bath
25. Dohány Street Synagogue
26. Rumbach Street Synagogue
27. Museum of Applied Arts
28. Erkel Theatre
29. Café Gerbeaud
30. Great Market Hall
31. Liberty Statue
32. Memento Park
33. Margaret Island
34. Aeropark

These venues and their near by locations:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parliament	47.507121	19.045669	Parlament	47.507041	19.045658	Capitol Building
1	Parliament	47.507121	19.045669	Kossuth Lajos tér	47.506896	19.047107	Plaza
2	Parliament	47.507121	19.045669	1956. In Memorian Kossuth tér	47.506370	19.046642	History Museum
3	Parliament	47.507121	19.045669	Bortársaság	47.505649	19.048543	Wine Shop
4	Parliament	47.507121	19.045669	Madal Cafe - Espresso & Brew Bar	47.507349	19.049045	Coffee Shop

Counts by neighborhood:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Aeropark	3	3	3	3	3	3
Andrássy Avenue	100	100	100	100	100	100
Aquincum Military Amphitheatre	46	46	46	46	46	46
Buda Castle	40	40	40	40	40	40
Café Gerbeaud	100	100	100	100	100	100
Castle Hill Funicular	54	54	54	54	54	54
City Park	29	29	29	29	29	29
Dohány Street Synagogue	100	100	100	100	100	100
Erkel Theatre	33	33	33	33	33	33
Fisherman's Bastion	95	95	95	95	95	95
Great Market Hall	100	100	100	100	100	100
Gresham Palace	100	100	100	100	100	100
Heroes' Square	64	64	64	64	64	64
House of Terror Museum	100	100	100	100	100	100
Hungarian Academy of Sciences	98	98	98	98	98	98
Hungarian National Gallery	39	39	39	39	39	39

There are around 220 unique venue categories.

The next step, analysis of each surounding famous location point was made:

```
----Buda Castle----
            venue  freq
0            Park  0.10
1  Hungarian Restaurant  0.08
2             Pub  0.08
3    Historic Site  0.08
4           Plaza  0.08


----Café Gerbeaud----
                    venue  freq
0                   Hotel  0.08
1              Restaurant  0.07
2                   Plaza  0.05
3        Italian Restaurant  0.04
4  Eastern European Restaurant  0.04


----Castle Hill Funicular----
            venue  freq
0  Hungarian Restaurant  0.09
```

1	Pub	0.07
2	Café	0.07
3	Hotel	0.06
4	Historic Site	0.06

From this infomation, the top 10 common venue of each tourist location was made:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Aeropark	Taxi Stand	Science Museum	Rest Area	Zoo Exhibit	Exhibit	French Restaurant	Fountain	Food Truck	Food & Drink Shop	Flower Shop
1	Andrássy Avenue	Coffee Shop	Pub	Café	Pizza Place	Bar	Gym	Italian Restaurant	Hostel	Hotel	Plaza
2	Aquincum Military Amphitheatre	Hungarian Restaurant	Indian Restaurant	Restaurant	Gym / Fitness Center	Gastropub	Food & Drink Shop	Nightclub	Chinese Restaurant	Eastern European Restaurant	Bar
3	Buda Castle	Park	Hungarian Restaurant	Pub	Plaza	Scenic Lookout	Café	Historic Site	Restaurant	Coffee Shop	Hotel
4	Café Gerbeaud	Hotel	Restaurant	Plaza	Eastern European Restaurant	Italian Restaurant	Ice Cream Shop	Coffee Shop	Theater	Dessert Shop	Café
5	Castle Hill Funicular	Hungarian Restaurant	Pub	Café	Plaza	Scenic Lookout	Historic Site	Hotel	Wine Bar	French Restaurant	Indie Theater
6	City Park	Park	Beer Garden	Bus Stop	Brewery	Café	Spa	Monument / Landmark	Castle	Gastropub	Skating Rink
7	Dohány Street Synagogue	Coffee Shop	Bar	Café	Breakfast Spot	Hostel	Bakery	Pub	Italian Restaurant	Thai Restaurant	Ice Cream Shop
8	Erkel Theatre	Hotel	Bakery	Bar	Irish Pub	Tattoo Parlor	Sushi Restaurant	Supermarket	Café	Social Club	Coffee Shop

4. Cluster Neighborhoods

After testing with K-elbow method, the dataset was process with 7 clusters:

```
# set number of clusters
kclusters = 7

budapest_grouped_clustering = budapest_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(budapest_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```
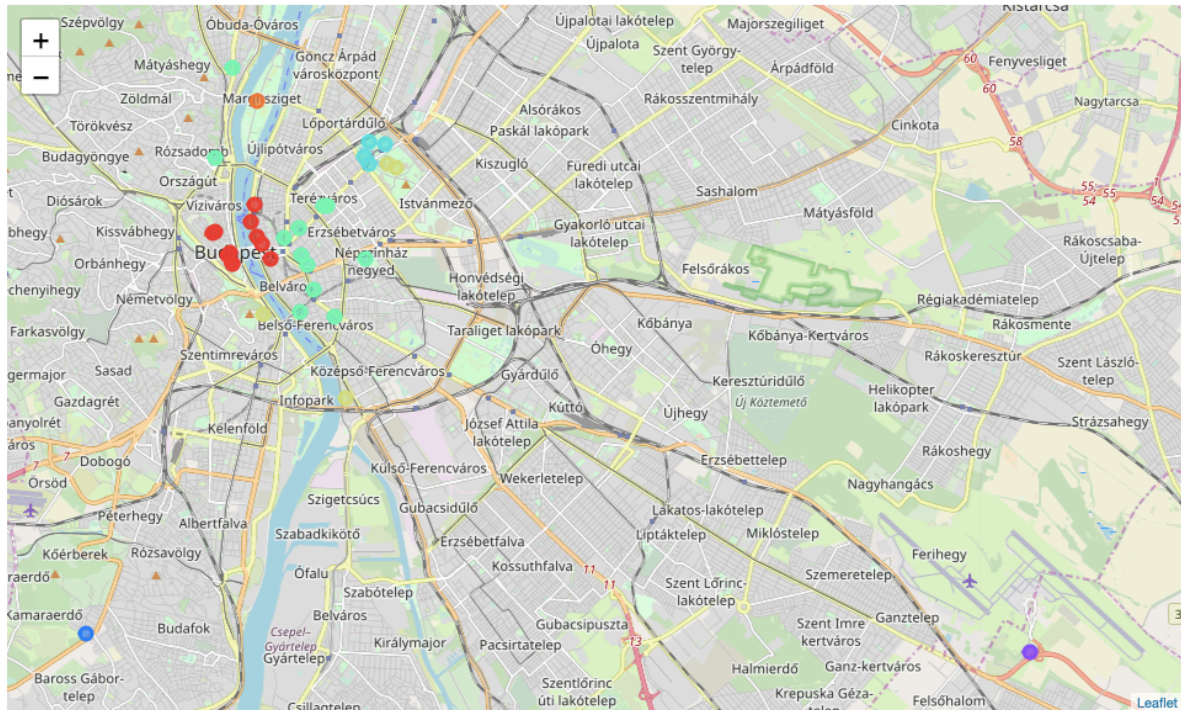
Add cluster labels and merge the name of location with its latitude, longitude:

	Name	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Parliament	47.507121	19.045669	2	Outdoor Sculpture	Italian Restaurant	Hungarian Restaurant	Bank	Café	French Restaurant	Capitol Building	Noodle House	Boat or Ferry
1	St. Stephen's Basilica	47.500890	19.053989	2	Hungarian Restaurant	Italian Restaurant	Hotel	Restaurant	Coffee Shop	Dessert Shop	Bar	Wine Bar	Cocktail Bar
2	Hungarian National Museum	47.491198	19.062476	2	Coffee Shop	Hotel	Italian Restaurant	Ice Cream Shop	Café	Bar	Chinese Restaurant	Restaurant	Pizza Place
3	National Theatre	47.470798	19.071066	0	Sandwich Place	Café	Concert Hall	Park	Music Venue	Bus Stop	Bakery	Candy Store	Skating Rink
4	Buda Castle	47.496205	19.039567	4	Park	Hungarian Restaurant	Pub	Plaza	Scenic Lookout	Café	Historic Site	Restaurant	Coffee Shop

The clusters distribution:

5. Examine Clusters:

Cluster 0th:

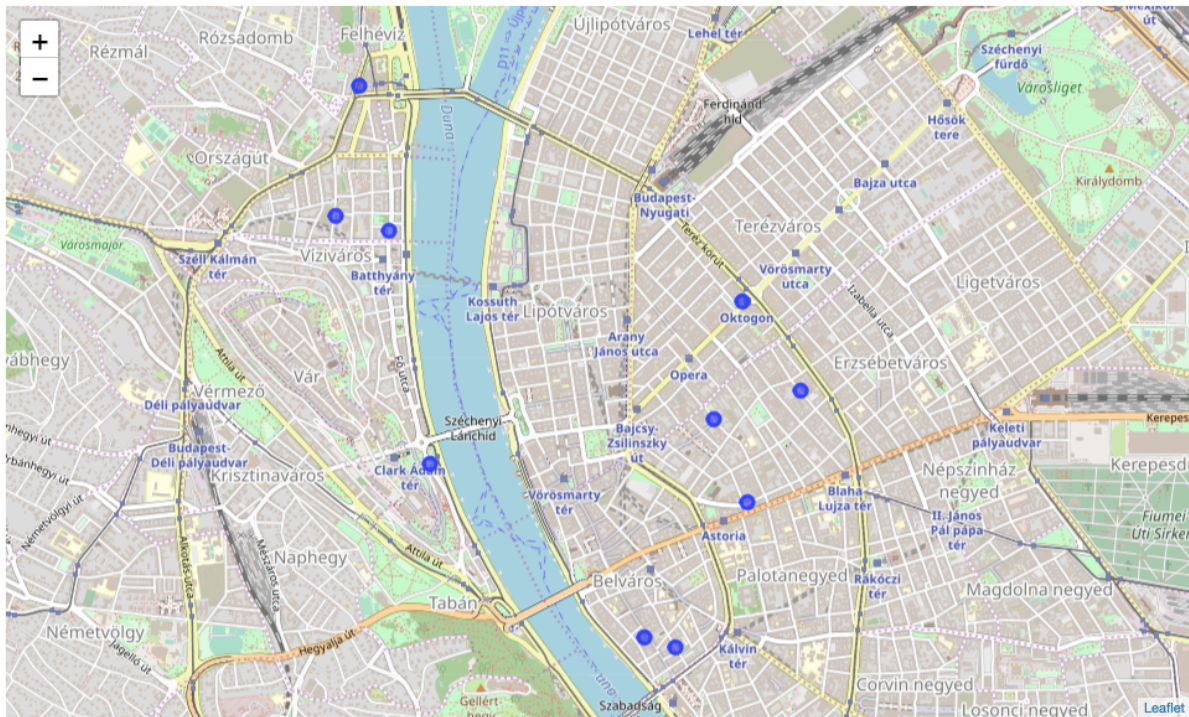|	Latitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	47.507121	Outdoor Sculpture	Italian Restaurant	Hungarian Restaurant	Bank	Café	French Restaurant	Capitol Building	Noodle House	Boat or Ferry	Mexican Restaurant
4	47.496205	Park	Hungarian Restaurant	Pub	Plaza	Scenic Lookout	Café	Historic Site	Restaurant	Coffee Shop	Hotel
5	47.496113	Park	Hungarian Restaurant	Café	Scenic Lookout	Historic Site	Pub	Plaza	Coffee Shop	Restaurant	Hotel
6	47.501954	Café	Hungarian Restaurant	Coffee Shop	Hotel	Restaurant	Pub	Historic Site	Plaza	Bar	Indie Theater
7	47.502183	Café	Hungarian Restaurant	Hotel	Restaurant	Coffee Shop	Pub	Bar	Plaza	Historic Site	Indie Theater
8	47.498191	Hungarian Restaurant	Pub	Café	Plaza	Scenic Lookout	Historic Site	Hotel	Wine Bar	French Restaurant	Indie Theater
9	47.497735	Hungarian Restaurant	Café	Pub	Plaza	Scenic Lookout	Hotel	Coffee Shop	Historic Site	Bistro	Restaurant
11	47.501246	Restaurant	Italian Restaurant	Hotel	Coffee Shop	Hungarian Restaurant	Plaza	Park	Eastern European Restaurant	Bistro	Dessert Shop
12	47.503970	Coffee Shop	Plaza	Italian Restaurant	Hungarian Restaurant	Hotel	Pub	Restaurant	Wine Bar	Park	Outdoor Sculpture
13	47.499713	Restaurant	Hotel	Italian Restaurant	Hungarian Restaurant	Coffee Shop	Eastern European Restaurant	Plaza	Dessert Shop	Bistro	Wine Bar
28	47.496933	Hotel	Restaurant	Plaza	Eastern European Restaurant	Italian Restaurant	Ice Cream Shop	Coffee Shop	Theater	Dessert Shop	Café

Cluster 4ᵗʰ:

	Latitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	47.500890	Hungarian Restaurant	Italian Restaurant	Hotel	Restaurant	Coffee Shop	Dessert Shop	Bar	Wine Bar	Cocktail Bar	Pizza Place
2	47.491198	Coffee Shop	Hotel	Italian Restaurant	Ice Cream Shop	Café	Bar	Chinese Restaurant	Restaurant	Pizza Place	Park
10	47.532942	Hungarian Restaurant	Indian Restaurant	Restaurant	Gym / Fitness Center	Gastropub	Food & Drink Shop	Nightclub	Chinese Restaurant	Eastern European Restaurant	Bar
14	47.515965	Yoga Studio	Café	Breakfast Spot	Coffee Shop	Bar	Spa	Nightclub	Park	Restaurant	Hotel
15	47.507102	Coffee Shop	Pub	Café	Pizza Place	Bar	Gym	Italian Restaurant	Hostel	Hotel	Plaza
16	47.502717	Bar	Theater	Hungarian Restaurant	Ice Cream Shop	Dessert Shop	Thai Restaurant	Restaurant	Café	Cocktail Bar	Coffee Shop
17	47.506893	Bar	Coffee Shop	Pub	Pizza Place	Theater	Indian Restaurant	Thai Restaurant	Burger Joint	Italian Restaurant	Café
24	47.495907	Coffee Shop	Bar	Café	Breakfast Spot	Hostel	Bakery	Pub	Italian Restaurant	Thai Restaurant	Ice Cream Shop
25	47.497705	Bar	Cocktail Bar	Café	Coffee Shop	Bakery	Hotel	Pizza Place	Hungarian Restaurant	Italian Restaurant	Breakfast Spot
26	47.486119	Bar	Pub	Hotel	Coffee Shop	Dessert Shop	Electronics Store	Italian Restaurant	Beer Garden	Café	Beer Bar
27	47.497103	Hotel	Bakery	Bar	Irish Pub	Tattoo Parlor	Sushi Restaurant	Supermarket	Café	Social Club	Coffee Shop
29	47.487085	Coffee Shop	Hungarian Restaurant	Bar	Hotel	Plaza	Café	Ice Cream Shop	Thai Restaurant	Bakery	Chinese Restaurant

Some more analysis:

Looking for milk-tea shop in Budapest center:

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng
0	Flying Bird Tea House	Tea Room	Kazinczy utca 52.	HU	Budapest	Magyarország	NaN	1521	[Budapest, Kazinczy utca 52., 1072, Magyarország]	[{'label': 'display', 'lat': 47.49945058120843...	47.499451	19.060641
1	Zhao Zhou Teashop & Lab	Tea Room	Lánchíd u. 5.	HU	Budapest	Magyarország	NaN	121	[Budapest, Lánchíd u. 5., Magyarország]	[{'label': 'display', 'lat': 47.49735359915229...	47.497354	19.041026
2	Buubbles Tea Company	Tea Room	Veres Pálné 33	HU	Budapest	Magyarország	NaN	1695	[Budapest, Veres Pálné 33, 1054, Magyarország]	[{'label': 'display', 'lat': 47.48879624688007...	47.488796	19.057986
3	Luis Tea Shop & Gourmet	Deli / Bodega	Andrássy Út 46	HU	Budapest XIII. kerülete	Magyarország	NaN	1822	[Budapest XIII. kerülete, Andrássy Út 46, Magy...	[{'label': 'display', 'lat': 47.50498175055748...	47.504982	19.062645
4	K29 Kávézó és Tea Szalon	Café	Kertész u. 29.	HU	Budapest	Magyarország	NaN	1989	[Budapest, Kertész u. 29., 1073, Magyarország]	[{'label': 'display', 'lat': 47.5008114873742,...	47.500811	19.066681
5	1000TEA	Tea Room	Váci u. 65.	HU	Budapest	Magyarország	NaN	1540	[Budapest, Váci u. 65., 1056, Magyarország]	[{'label': 'display', 'lat': 47.48925378899116...	47.489254	19.055868
6	Milkshake Beauty	Health & Beauty Service	NaN	HU	NaN	Magyarország	NaN	1115	[Magyarország]	[{'label': 'display', 'lat': 47.50828290393077...	47.508283	19.038189
7	Kávé - tea	Bar	Síp Utca 2	HU	Budapest	Magyarország	Dohány Utca	1721	[Budapest, Síp Utca 2 (Dohány Utca), 1074, Mag...	[{'label': 'display', 'lat': 47.4955676836366,...	47.495568	19.062979
8	Milka kvártély	None	NaN	HU	NaN	Magyarország	NaN	1886	[Magyarország]	[{'label': 'display', 'lat': 47.515078, 'lng':...	47.515078	19.036162
9	Kávé Tea Rum Bum	Café	NaN	HU	NaN	Magyarország	NaN	1261	[Magyarország]	[{'label': 'display', 'lat': 47.508963, 'lng':...	47.508963	19.034462

Reveal them in the map

6. Results

- After all, we can can see clearly that cluster 0th and cluster 4th is the most potential place for me to start-up the business. But then there are must more most common coffee places in cluster 4th. So I will make the decision to put the milk tea shop in cluster 0th area.
- And from the milk tea shop distribution map showed that the city center, near Deak Ferenc square, Arany Janos street can be the potential places for us to put the shop, because of there was no milk tea shop in this area.

7. Discussion

While doing descriptive analysis we found some interesting point:

- In general, this result can be effected by a lot of different real-life factors. But this is a good base for the stockholder to make the decision.
- All of the data, API which is used in this project was free. Better data can be collect to give the better result.

- This project helps me a lot to to understand about how to deal with the data, and using Foursquare API to access valuable data. This will be the great support for me to study with data.

8. Conclusion

- Milk tea shops is rare in Budapest, which is the capital of Hungary, where attract millions of visitors each year. This is the huge potential profit if we can take the advantage of it.
- Since milk tea is well-known in Asia, so the targeted customer is Asian tourist, then broadening the brand to EU is the next step.
- From this project, I realized Budapest is a enormous city and it gonna be important to learn customer behavior if I want to sell food or drink. Learning which taste the customer, giving some discount for student, and so on, with the help of Data Science are ways to solve this business problem.

9. Recomendation

The analysis can be improved with other extensions like:

- Consider more categories, like: Restaurant, trending, shopping places, ... In order to attract more customer
- Can use the cluster center to choose the best location in the area (but this can increase the renting coust of the shop, because it is only allows to put the shop in the small area

Thank you for your review. Give me more comment via email:

hoangleminhtien@gmail.com