Selecting place to open a bar in Ho Chi Minh city

Table of contents

1. Introduction

My friends are planning to start new business in Ho Chi Minh city. They would like to open a new bar, and they are looking a good area to open it. This capstone project is helping them on making decision.

The areas we will target to:

- as close to center of each neighborhood as possible, which more potential customers
- not too many bars already existed, which as less competitors as possible

2. Data acquisition and cleaning

2.1 Data Sources

To analyze place for a new bar opening, I would like to use

- First of all, list of boroughs and neighborhoods of Ho Chi Minh City. They are available from General Statistics Office of Vietnam website https://www.gso.gov.vn/dmhc2015/Default.aspx. This page provides button to export all boroughs and neighborhoods of Ho Chi Minh City. After exporting, data is stored to a file and uploaded to a cloud service which is easily accessed from the notebook
- Second, I would use Google Maps API to get lat/long of each neighborhood before exploring them with Foursquare API
- Next, Foursquare API helps to get bars and similar places which people usually visit and have a good feedback
- Finally, I would use power of k-means clustering algorithm to analyse on these places

2.2 Preprocessing Ho Chi Minh City data

Download boroughs and neighborhoods of Ho Chi Minh City from the page https://www.gso.gov.vn/dmhc2015/Default.aspx. Data is already in csv format, then I have saved it to Google Drive.

To get geolocation of these neighborhoods, Google Maps Platform APIs should be used. Because the address name transfering to Google Map APIs is Vietnamese, I would like to assign some meaningful column headers

to following variables:

- cityname
- boroughname
- neighborhoodname

Thefore, the address name would be the form of neighborhoodname + boroughname + cityname

Lastly, using Google API to get lat long of all these places.

```
Coordinate of Phường Tân Định, Quận 1, Thành phố Hồ Chí Minh: [10.7930968, 106.6902951, 'Tan Dinh, District 1, Ho Chi Minh City, Vietnam']
Coordinate of Phường Đa Kao, Quận 1, Thành phố Hồ Chí Minh: [10.7878843, 106.6984026, 'Da Kao, District 1, Ho Chi Minh City, Vietnam']
Coordinate of Phường Bến Nghé, Quận 1, Thành phố Hồ Chí Minh: [10.7808334, 106.702825, 'Bến Nghé, District 1, Ho Chi Minh City, Vietnam']
Coordinate of Phường Bến Thành, Quận 1, Thành phố Hồ Chí Minh: [10.7744331, 106.6954544, 'Ben Thanh, District 1, Ho Chi Minh City, Vietnam']
Coordinate of Phường Nguyễn Thái Bình, Quận 1, Thành phố Hồ Chí Minh: [10.7693846, 106.7006138, 'Nguyen Thai Binh, District 1, Ho Chi Minh City, Vietnam']
Coordinate of Phường Phạm Ngũ Lão, Quận 1, Thành phố Hồ Chí Minh: [10.7658855, 106.6908105, 'Pham Ngu Lao, District 1, Ho Chi Minh City, Vietnam']
Coordinate of Phường Cầu Ông Lãnh, Quận 1, Thành phố Hồ Chí Minh: [10.7655446, 106.6961914, 'Cầu Ông Lãnh, District 1, Ho Chi Minh City, Vietnam']
Coordinate of Phường Cô Giang, Quận 1, Thành phố Hồ Chí Minh: [10.7616235, 106.6932433, 'Co Giang, District 1, Ho Chi Minh City, Vietnam']
Coordinate of Phường Nguyễn Cư Trinh, Quận 1, Thành phố Hồ Chí Minh: [10.7640301, 106.68661, 'Nguyen Cu Trinh, District 1, Ho Chi Minh City, Vietnam']
Coordinate of Phường Cầu Kho, Quận 1, Thành phố Hồ Chí Minh: [10.7577834, 106.6888211, 'Cau Kho, District 1, Ho Chi Minh City, Vietnam']
Coordinate of Phường Thạnh Xuân, Quận 12, Thành phố Hồ Chí Minh: [10.8834303, 106.6703963, 'Thạnh Xuân, District 12, Ho Chi Minh City, Vietnam']
Coordinate of Phường Thạnh Lộc, Quận 12, Thành phố Hồ Chí Minh: [10.8712302, 106.6859815, 'Thạnh Lộc, Quận 12, Thành phố Hồ Chí Minh, Vietnam']
Coordinate of Phường Hiệp Thành, Quận 12, Thành phố Hồ Chí Minh: [10.8825023, 106.6379724, 'Hiệp Thành, District 12, Ho Chi Minh City, Vietnam']
Coordinate of Phường Thới An, Quận 12, Thành phố Hồ Chí Minh: [10.8760697, 106.6556575, 'Thoi An, District 12, Ho Chi Minh City, Vietnam']
Coordinate of Phường Tân Chánh Hiệp, Quận 12, Thành phố Hồ Chí Minh: [10.866797, 106.6261831, 'Tân Chánh Hiệp, Quận 12, Thành phố Hồ Chí Minh, Vietnam']
```

2.3 Preprocessing Foursquare data

Foursquare API provides great information about places. In this project, I would explore top 100 places within the neighborhood center using Foursquare APIs. Places nearby center of each neighborhood is expecting to have crowded people, it is good to have a lot informatoin aroud these places.

```
Tan Dinh, District 1, Ho Chi Minh City, Vietnam
Da Kao, District 1, Ho Chi Minh City, Vietnam
Bến Nghé, District 1, Ho Chi Minh City, Vietnam
Ben Thanh, District 1, Ho Chi Minh City, Vietnam
Nguyen Thai Binh, District 1, Ho Chi Minh City, Vietnam
Pham Ngu Lao, District 1, Ho Chi Minh City, Vietnam
Cầu Ông Lãnh, District 1, Ho Chi Minh City, Vietnam
Co Giang, District 1, Ho Chi Minh City, Vietnam
Nguyen Cu Trinh, District 1, Ho Chi Minh City, Vietnam
Cau Kho, District 1, Ho Chi Minh City, Vietnam
Thạnh Xuân, District 12, Ho Chi Minh City, Vietnam
Thạnh Lộc, Quận 12, Thành phố Hồ Chí Minh, Vietnam
Hiệp Thành, District 12, Ho Chi Minh City, Vietnam
Thoi An, District 12, Ho Chi Minh City, Vietnam
Tân Chánh Hiệp, Quận 12, Thành phố Hồ Chí Minh, Vietnam
An Phú Đông, Quận 12, Thành phố Hồ Chí Minh, Vietnam
Tân Thới Hiệp, District 12, Ho Chi Minh City, Vietnam
Trung My Tay, District 12, Ho Chi Minh City, Vietnam
Tân Hưng Thuận, Quận 12, Thành phố Hồ Chí Minh, Vietnam
Đông Hưng Thuận, Quận 12, Thành phố Hồ Chí Minh, Vietnam
Tan Thoi Nhat, District 12, Ho Chi Minh City, Vietnam
Linh Xuan, Thủ Đức, Ho Chi Minh City, Vietnam
Binh Chieu, Thủ Đức, Ho Chi Minh City, Vietnam
Linh Trung, Thủ Đức, Ho Chi Minh City, Vietnam
Tam Bình, Thủ Đức, Ho Chi Minh City, Vietnam
Tam Phú, Thủ Đức, Ho Chi Minh City, Vietnam
Hiep Binh Phuoc, Thủ Đức, Ho Chi Minh City, Vietnam
Hiep Binh Chanh, Thủ Đức, Ho Chi Minh City, Vietnam
Linh Chiểu, Thủ Đức, Thành phố Hồ Chí Minh, Vietnam
```

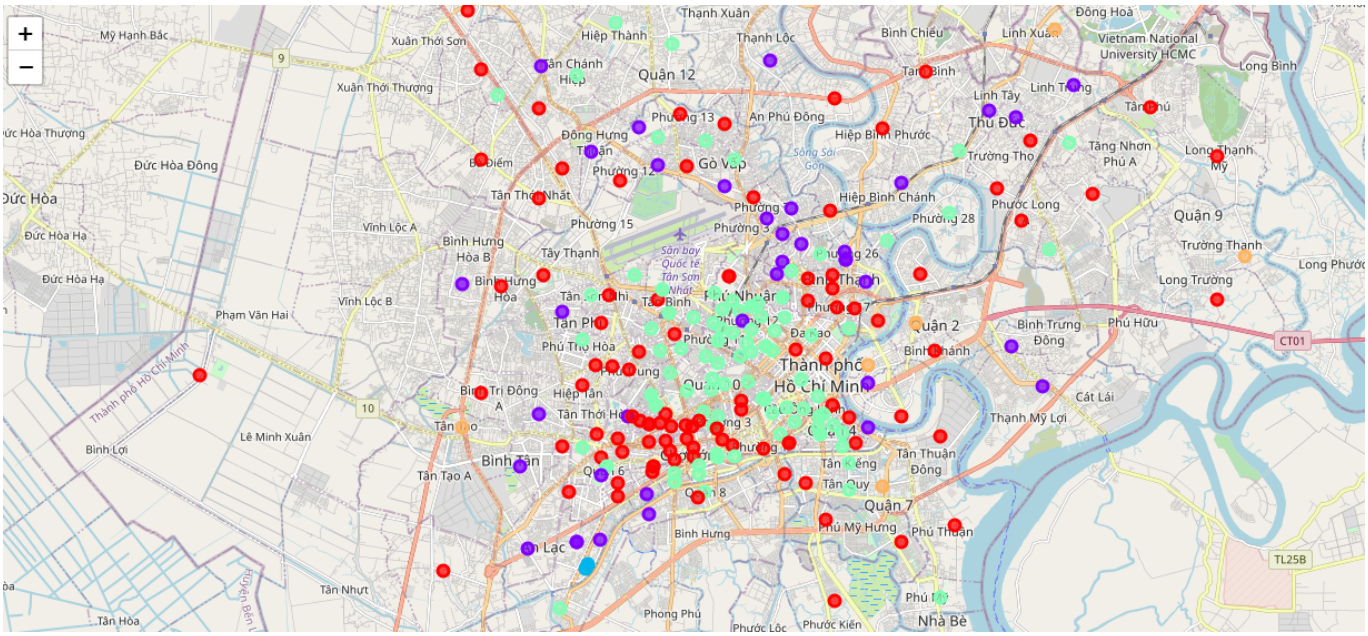
Top visited places of each neighborhoods

	Neighbourhood	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	01 Cao Thắng, Phường 2, Quận 3, Thành phố Hồ C...	Vietnamese Restaurant	Café	Asian Restaurant	Bookstore	Bakery	Noodle House	Chinese Restaurant	Food Truck	Pizza Place	Hotel
1	09 Cao Thắng, Phường 2, Quận 3, Thành phố Hồ C...	Vietnamese Restaurant	Café	Asian Restaurant	Bakery	Bookstore	Food Truck	Noodle House	Chinese Restaurant	Seafood Restaurant	Vegetarian / Vegan Restaurant
2	107 Cao Văn Lầu, Phường 1, Quận 6, Thành phố H...	Food	Food Truck	Market	Dessert Shop	Yoga Studio	Dive Bar	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store
3	137 Nguyễn Văn Đậu, Phường 7, Bình Thạnh, Thàn...	Café	Vietnamese Restaurant	Asian Restaurant	Brewery	Stadium	Yoga Studio	Diner	Fast Food Restaurant	Fabric Shop	Exhibit
4	155, Đường Nguyễn Văn Trỗi, Phường 11, Phú Nhu...	Café	Coffee Shop	Vietnamese Restaurant	Hotel	Chinese Restaurant	Juice Bar	BBQ Joint	Flea Market	Bed & Breakfast	Bar

3. Methodology

To have better insight of places around neighborhood, I would use k-means algorithm to cluster each area, then count total existed bars for each cluster Before applying k-means, data have to be grouped and transformed using one hot encoding technique.

K-means then is used to cluster popuplar places in Ho Chi Minh City



4. Result and Discussion

Total visited bars in each cluster:

```
clusters = []
for i in range(kclusters):
  print("Checking on cluster %d" % (i))
  clusters.append(hcmc_merged_nona.loc[hcmc_merged_nona['Cluster Labels'] == i, hcmc_merged_nona.columns[[2] + list(range(4, hcmc_merged_nona.shape[1]))]])

popular_bars_per_cluster = []
for i in range(kclusters):
  # print("shape of cluster %d: %s" % (i, clusters[i].shape))
  total = 0
  for col in clusters[i]:
    count = clusters[i][col].str.count("[bB]ar").sum()
    if (count > 0):
      total += count
      # print("found %d from %s" % (count, col))
  popular_bars_per_cluster.append(total)
  print("number of popular bars on cluster %d: %d" % (i, popular_bars_per_cluster[i]))
```

```
Checking on cluster 0
Checking on cluster 1
Checking on cluster 2
Checking on cluster 3
Checking on cluster 4
number of popular bars on cluster 0: 7
number of popular bars on cluster 1: 25
number of popular bars on cluster 2: 2
number of popular bars on cluster 3: 21
number of popular bars on cluster 4: 0
```

From the clustering result above, we have seen that cluster 0 and 3 have most bars, therefore it will have more competitors to open new bars in these areas. Howver, other cluster 1, 2 and 4 have a few bars, less competitors on these areas

Result of each cluster:

- Cluster 0

```
clusters[0]
```

95	Phường 14, Tân Bình, Thành phố Hồ Chí Minh, Vi...	Asian Restaurant	Yoga Studio	Dim Sum Restaurant	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
101	Tan Thanh, Tân Phú, Ho Chi Minh City, Vietnam	Asian Restaurant	Yoga Studio	Dim Sum Restaurant	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
103	Phu Thanh, Tân Phú, Ho Chi Minh City, Vietnam	Hotpot Restaurant	Asian Restaurant	Juice Bar	Furniture / Home Store	Yoga Studio	Diner	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit
131	480 Nguyễn Thị Định, Phường Thạnh Mỹ Lợi, Quận...	Asian Restaurant	Café	Yoga Studio	Diner	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant
170	Phường 8, District 11, Ho Chi Minh City, Vietnam	Basketball Stadium	Asian Restaurant	Department Store	Café	Yoga Studio	Dive Bar	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit
171	Phường 9, District 11, Ho Chi Minh City, Vietnam	Pet Store	Asian Restaurant	Department Store	Café	Basketball Stadium	Yoga Studio	Dive Bar	Fish Market	Fast Food Restaurant	Fabric Shop
176	Phường 1, District 11, Ho Chi Minh City, Vietnam	Vegetarian / Vegan Restaurant	Asian Restaurant	Noodle House	Yoga Studio	Dim Sum Restaurant	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant

- Cluster 1

```
clusters[1]
```

	Vie...			Restaurant		Restaurant			Store	Restaurant	Restaurant
243	Bình Trị Đông A, Bình Tân, Thành phố Hồ Chí Mi...	Shoe Store	Yoga Studio	Dim Sum Restaurant	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
249	Tan Thuan Dong, District 7, Ho Chi Minh City, ...	Gay Bar	Yoga Studio	Dim Sum Restaurant	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
255	Phú Thuận, District 7, Ho Chi Minh City, Vietnam	Diner	Yoga Studio	Flea Market	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
256	Tân Phú, District 7, Ho Chi Minh City, Vietnam	Hotel	Italian Restaurant	Convenience Store	Breakfast Spot	Sports Bar	Athletics & Sports	Coffee Shop	Greek Restaurant	Gym / Fitness Center	Hardware Store
257	Tân Phong, District 7, Ho Chi Minh City, Vietnam	Indian Restaurant	Korean Restaurant	Bubble Tea Shop	Hotel	Bar	Gastropub	Japanese Restaurant	Fast Food Restaurant	Multiplex	Dessert Shop
280	Hoc Mon, Hóc Môn, Ho Chi Minh City, Vietnam	Restaurant	Vegetarian / Vegan Restaurant	Food & Drink Shop	Café	Market	Seafood Restaurant	Yoga Studio	Dim Sum Restaurant	Exhibit	Electronics Store
282	Nhị Bình, Hóc Môn, Ho Chi Minh City, Vietnam	Cafeteria	Yoga Studio	Dim Sum Restaurant	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant	Dumpling Restaurant

- Cluster 2

```
clusters[2]
```

186	Phường 18, District 4, Ho Chi Minh City, Vietnam	Boat or Ferry	Café	Yoga Studio	Diner	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant
215	Phường 11, District 6, Ho Chi Minh City, Vietnam	Café	Supermarket	Vietnamese Restaurant	Dim Sum Restaurant	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant
230	Phường 16, District 8, Ho Chi Minh City, Vietnam	Baseball Stadium	Café	Cafeteria	Yoga Studio	Diner	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store
234	Phường 14, District 8, Ho Chi Minh City, Vietnam	Chinese Restaurant	Café	Yoga Studio	Diner	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant
235	Phường 6, District 8, Ho Chi Minh City, Vietnam	Fast Food Restaurant	Café	Yoga Studio	Dim Sum Restaurant	Fish Market	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
236	Phường 15, District 8, Ho Chi Minh City, Vietnam	Miscellaneous Shop	Café	Yoga Studio	Diner	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant

- Cluster 3

```
clusters[3]
```

195	259/9 Nguyễn Tri Phương, Phường 5, Quận 10, Th...	Vietnamese Restaurant	Coffee Shop	Park	Noodle House	Dessert Shop	Seafood Restaurant	Market	Café	Ice Cream Shop	Asian Restaurant
198	2 Nguyễn Tri Phương, Phường 6, Quận 11, Thành ...	Coffee Shop	Dim Sum Restaurant	Vietnamese Restaurant	Seafood Restaurant	Dumpling Restaurant	Dessert Shop	BBQ Joint	Chinese Restaurant	Cruise Ship	Dive Bar
205	Phường 5, District 5, Ho Chi Minh City, Vietnam	Vietnamese Restaurant	Hotel	Flea Market	Chinese Restaurant	Vegetarian / Vegan Restaurant	Movie Theater	Dumpling Restaurant	Noodle House	Dessert Shop	Diner
210	Phường 13, District 6, Ho Chi Minh City, Vietnam	Flea Market	Fast Food Restaurant	Movie Theater	Vietnamese Restaurant	Food & Drink Shop	Asian Restaurant	Diner	Fabric Shop	Exhibit	Electronics Store
211	District 6, Ho Chi Minh City, Vietnam	Vietnamese Restaurant	Asian Restaurant	Diner	Yoga Studio	Dim Sum Restaurant	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store
212	District 6, Ho Chi Minh City, Vietnam	Vietnamese Restaurant	Asian Restaurant	Diner	Yoga Studio	Dim Sum Restaurant	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store
214	District 6, Ho Chi Minh City, Vietnam	Vietnamese Restaurant	Asian Restaurant	Diner	Yoga Studio	Dim Sum Restaurant	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store

- Cluster 4

```
clusters[4]
```

	fmt_name	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
21	Linh Xuan, Thủ Đức, Ho Chi Minh City, Vietnam	Vietnamese Restaurant	Yoga Studio	Dim Sum Restaurant	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
41	Trường Thạnh, Quận 9, Thành phố Hồ Chí Minh, V...	Vietnamese Restaurant	Furniture / Home Store	Yoga Studio	Diner	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant
125	Binh An, District 2, Ho Chi Minh City, Vietnam	Vietnamese Restaurant	Badminton Court	Yoga Studio	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
129	An Khánh, District 2, Ho Chi Minh City, Vietnam	Vietnamese Restaurant	Yoga Studio	Dim Sum Restaurant	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
245	Tan Tao, Bình Tân, Ho Chi Minh City, Vietnam	Vietnamese Restaurant	Yoga Studio	Dim Sum Restaurant	Fish Market	Fast Food Restaurant	Fabric Shop	Exhibit	Electronics Store	Eastern European Restaurant	Dumpling Restaurant

5. Conclusion

Despite, we have already known that cluster 0 and 3 will have more struggle to open bars but they also have a lot of customers because most of people visit bars in these areas Therefore, it depends on the business owner, how they would their business to be. If they would expect less competitors, they should open bars in area of cluster 1, 2 or 4. However, if they would expect more competitors with more potential customers, they would open bars in area of cluster 0 or 3