**Final report**
**Coursera Capstone – The Battles of Neighborhoods**
**Recommendation for the set of touris office place**

## 1. Introduction

I would like to introduce my problem related to opening the tourist office around Hanoi, Vietnam.

In Hanoi, there are a significant number of individual foreign tourism who do not book a completed tour before their travel. The tourist market is lacking the cheap quality mini trip service such as around Hanoi on bike, food tour or walking tour. I and my college plan to run the business in this sector and we are looking for the place to open 4-5 offices around Hanoi to attract the foreign customer.

My problem is where are the good places to open these tourist service office?

The requirement of the office area should be:

- There are many foreign individual tourists around (they could live, have food, drink near there)
- The distance between the office should be optimal to gather as much as the customer all the office can.

## 2. Data

To do this project, I need the data about the location of the place I have mentioned above which are

- The hotel
- The bar, club
- The food court, restaurant
- The tourist place (museum, walking street, ...)

To find these place, firstly, I will find the top 10 best hotel in term of price, quality and so on by Tripadvisor website.

*Table 1: Best hotels with latitude and longitude dataframe*

| | hotel | Latitude | Longitude |
|---|---|---|---|
| 0 | Hanoi La Siesta Hotel & Spa | 21.034234 | 105.853225 |
| 1 | O'Gallery Premier Hotel & Spa | 21.029668 | 105.845626 |
| 2 | Golden Sun Suites Hotel | 21.032632 | 105.849347 |
| 3 | Khách sạn Hà Nội La Siesta Diamond | 21.031594 | 105.854946 |
| 4 | Serene Boutique Hotel & Spa | 21.035007 | 105.847460 |

Then I will use Foursquare to find these place around the hotels. The dataframe of these place includes 1468 places.

*Table 2: Travel place around the hotel dataframe wil the number of row*

| | Venue | Venue_lat | Venue_lng |
|---|---|---|---|
| 0 | Bun Cha Ta | 21.034373 | 105.854382 |
| 1 | Orchid Cooking Class & Restaurant | 21.033874 | 105.853270 |
| 2 | Bami Bread (Bánh Mì Bami) | 21.034072 | 105.851321 |
| 3 | Phở Sướng | 21.033518 | 105.852039 |
| 4 | Gia Ngu Restaurant | 21.033029 | 105.852704 |

```
print (len(place))
```

```
1468
```

Here the map of all the places in the above dataframe.

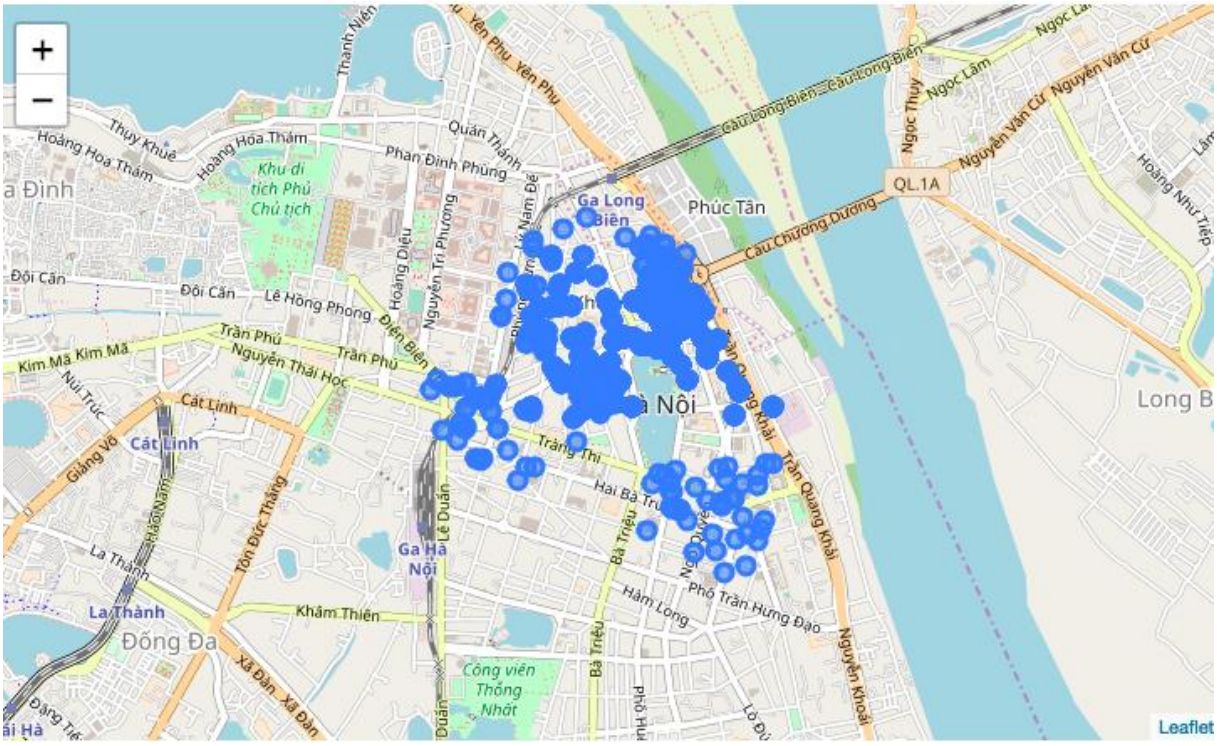


*Figure 1: Map of all the travel places*

## 3. Methodology

For this problem, I use K-mean clustering to classify the place into sub areas. I plan to open 5 office in Hanoi, therefore I used the k parameter of 5. Here is the dataframe of the place with the label haved been clusted. Label is from 0 to 4.

*Table 3: Travel places with labels dataframe*

| | Venue | Venue_lat | Venue_lng | labels |
|---|---|---|---|---|
| 0 | Bun Cha Ta | 21.034373 | 105.854382 | 1 |
| 1 | Orchid Cooking Class & Restaurant | 21.033874 | 105.853270 | 1 |
| 2 | Bami Bread (Bánh Mì Bami) | 21.034072 | 105.851321 | 1 |
| 3 | Phở Sướng | 21.033518 | 105.852039 | 1 |
| 4 | Gia Ngu Restaurant | 21.033029 | 105.852704 | 1 |

Below is the map of all the place which have been classified into 5 sub area with different colors.

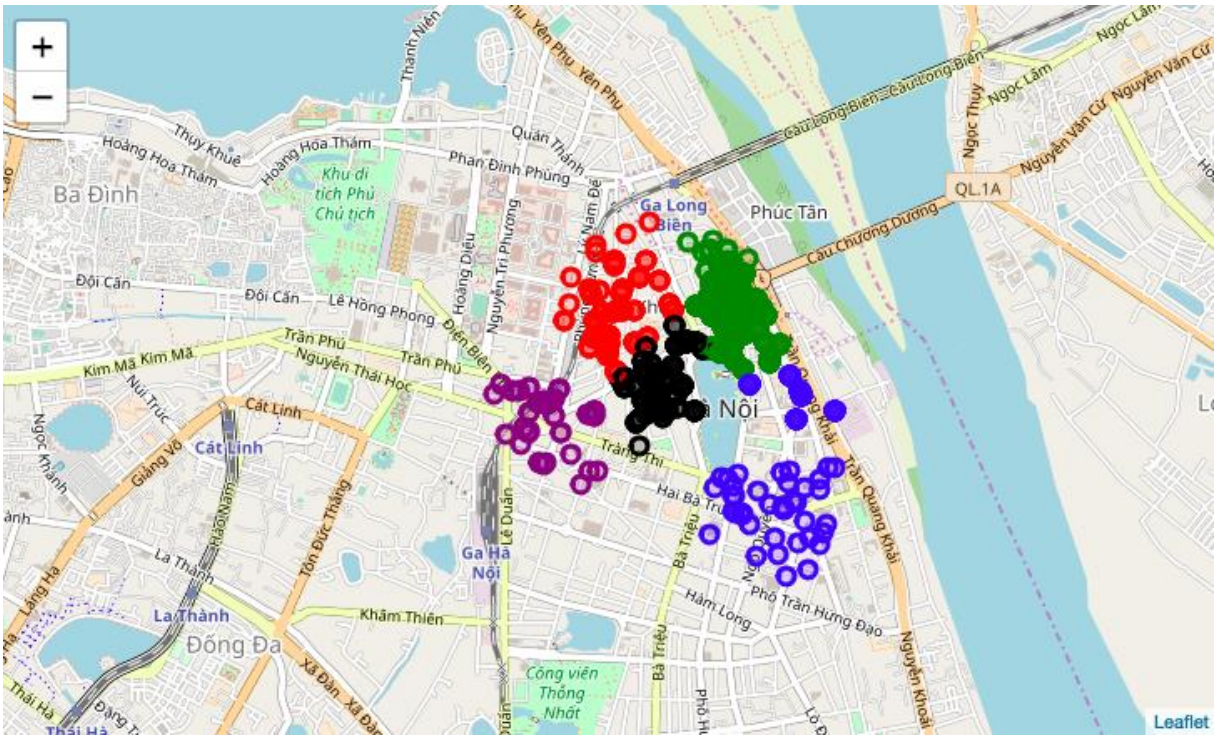


*Figure 2: Map of the clustered travel places*

## 4. Result

Here comes the result: I used k = 5 for the k-means clustering and from the above map there are quite clear 5 area with high potential of foreign travelers living, drinking, having the meal around when they were staying at Hanoi.

I will try to find the places for the office in the center of each clustered area.

## 5. Discussion

The problem of this method is that I pretended the travel and food place according to the Foursquare query are the place for foreign traveler. Another problem is that the model did not care about other tourism office in the area. The competitive aspect is removed from the model because the lack of information and data.

## 6. Conclusion

In this project, through a k-means cluster algorithm I separate the tourism area in the Hanoi center into 5 sub areas. To get the data related to these area, I come from the information of best 10 hotels according to the Tripadvisor website.