LOCATION RECOMMENDATION FOR A SOUVENIR SHOP

Cao Thi Nham

March 06, 2020

1   INTRODUCTION

1.1   Background

Danang is Vietnam's third largest city, and is one of the most attractive location for tourists in Vietnam. As a citizen of this city, I decided to use Danang in my project.

The city of Danang is about 1,256 square kilometres in size and divided in to 8 districts. From the fact that there are more than 6 million arrivals to this city annually, opening a souvenir shop is potentially successful. From an investor's view, locations are an important factor to the susccess of the business. In fact, not all of the areas attract tourists equally, therefore I assume that locations that have high visitor density are potential places to do business of selling sourvenirs. However, it is difficult to acquire information that will guide investors in this direction.

1.2   Problem

Data about check-ins of venues in the city may contribute to determining locations that are suitable for the business of souvenir shop.

1.3   Interest

Obviously, investors would be very interested in the map that show potential locations for their business of selling a souvenir shop.

2   DATA ACQUISITION & CLEANING

2.1   Data source

To solve problem, I used the following data source:

- Data from Foursquare to get common venues and their numbers of visitors/checkins.

2.2   Data acquisition

To retrieve numbers of visitings of venues in Danang, firstly I used Foursquare API to get popular venues in Danang, and the result is 100 venues with details information as follow.

	id	name	categories	lat	lng	district
0	5aabf13ed552c7495c4b16cc	Thien Ly Danang-style	Vietnamese Restaurant	16.076115	108.218690	Đà Nẵng
1	5a26a41a31ac6c676705e94c	cộng cà phê	Coffee Shop	16.068063	108.223512	Đà Nẵng
2	50389430e4b0fc42f753b8a3	Luna Pub	Italian Restaurant	16.079725	108.222998	Đà Nẵng
3	582b1746641e5c276f013971	Pizza 4P's	Pizza Place	16.062675	108.222845	Đà Nẵng
4	59d6062a67af3a61ae531b42	Herbal Spa	Spa	16.069384	108.236756	Đà Nẵng

Secondly, I used Foursquare API again to get detail information of each venue, especially visitsCount. And the result as the sample data:

|	id	name	categories	lat	lng	district	visitsCount
0	5aabf13ed552c7495c4b16cc	Thien Ly Danang-style	Vietnamese Restaurant	16.076115	108.218690	Đà Nẵng	251
1	5a26a41a31ac6c676705e94c	cộng cà phê	Coffee Shop	16.068063	108.223512	Đà Nẵng	584
2	50389430e4b0fc42f753b8a3	Luna Pub	Italian Restaurant	16.079725	108.222998	Đà Nẵng	1635
3	582b1746641e5c276f013971	Pizza 4P's	Pizza Place	16.062675	108.222845	Đà Nẵng	1483
4	59d6062a67af3a61ae531b42	Herbal Spa	Spa	16.069384	108.236756	Đà Nẵng	377

Finally, I used folium python library to visualize the positions of venues on the map.
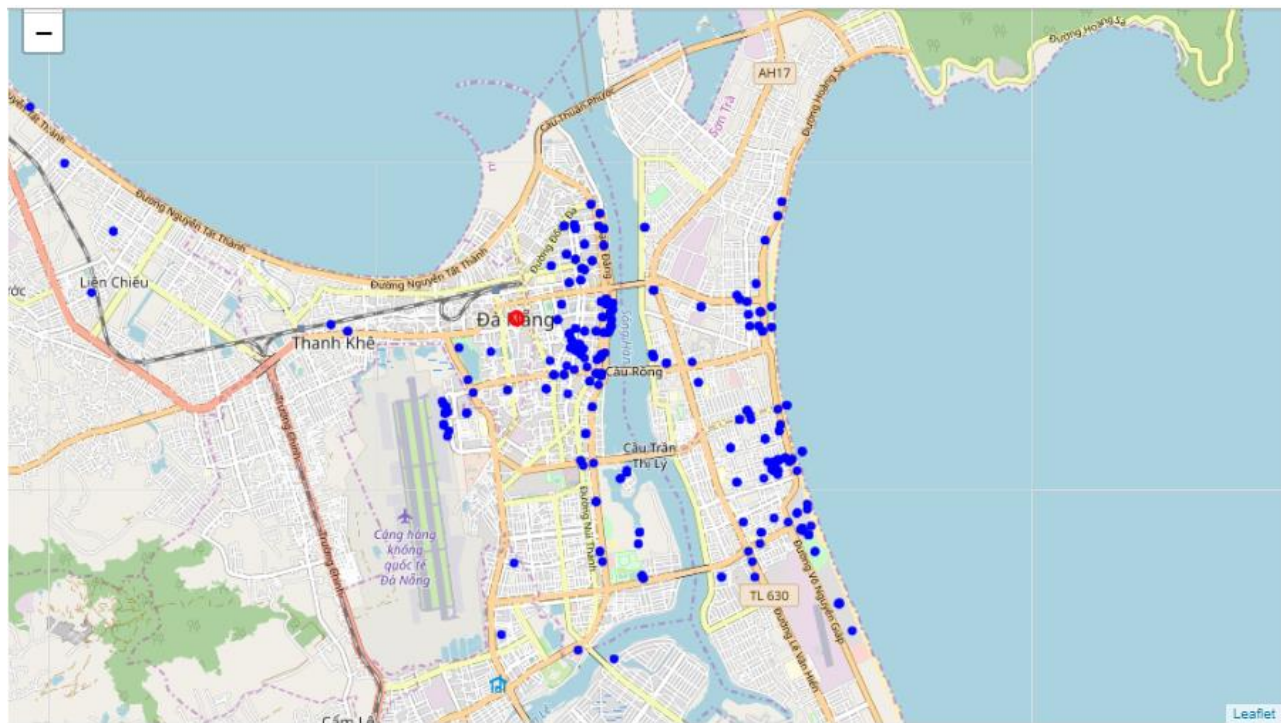
From the map, we can see that venues are quite dense in the center of the city.

3 METHODOLOGY
Firstly, I find the best k for k-mean algorithm by using Elbow method.
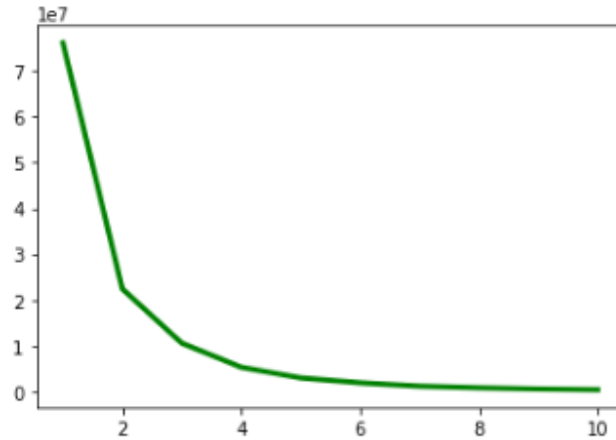
According to line graph, the optimal value for k is 2, thus I segment venues into 2 clusters based on visitsCount.

Next steps, I used k-mean to cluster venues.

```
kclusters=2

#run k-mean cluster
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(points)

# check cluster labels generated for each row in the dataframe
kmeans.labels_
```

```
array([[0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0,
        0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0])
```

Finally, I visualized the clusters.
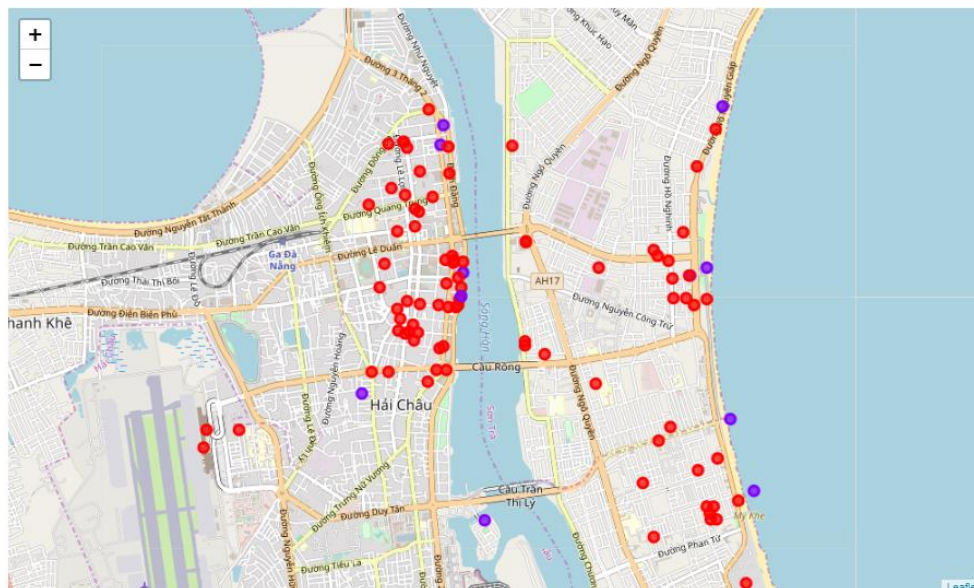
4   RESULTS

From the above map, red pots are venues with high numbers of visitors. As I assumed above, the higher visitor density, the more potential to start business with souvenir shop. Consequently, the area between Cầu Rồng brigde and Sông Hàn brigde is potential location.

5   CONCLUTION

As I mentioned above, Da Nang is a tourism city in Vietnam. It is very potential to open a souvenir shop to do business and introduce Vietnamese culture. In this project, I retrieved data of popular venues in Da Nang by using Foursquare API and then segmented them by numbers of visitings to find the optimal location to start the business. Based on the map, investors could clearly realize the locations they can start with.