Cluster District in Bangkok

Narongchai Jaroonpipatkul

May 05, 2021

1. Introduction

1.1 Background

In my hometown, Bangkok is the most visited city in the world, with comforts of megamalls, modern spaces, golden temples and shrines. Follow your nose to Bangkok's countless street markets stuffed with spicy, sour, sweet, salty flavors, and hop on a boat to reach their floating markets.

However, Bangkok is very high-density population and very high traffic jams, with approximately 22 million international visitors. There are 10.7 million people live and population density of 5,294 people per square kilometer, divided into 50 administrative districts.

If, I want to buy a new house or condo in Bangkok, it is advantageous for me to cluster each district in Bangkok for decision what location is best to buy.

1.2 Problem

Data include venues and population density might contribute to determining each location. So, this project aim to cluster each district in Bangkok according to population density and social place density.

1.3 Interest

Those who want to buy a new residence, want to know the information as much as possible. Buy a new residence is a big deal in life. It's better if have much information to support the decision.

2. Data acquisition and cleaning

2.1 Data sources

- I used Forsquare API to get the venues of each district in Bangkok.
- I scrape common districts in Bangkok data from wikipedia.
- I use my owned data about area in each district in Bangkok and calculate density population.

2.2 Data cleaning

After searching for data on many websites, some websites have data not match. So, I decide to select from the most reliable source.

There are not too many public datas related to demographic and social parameters. In many case, I setup my on table.

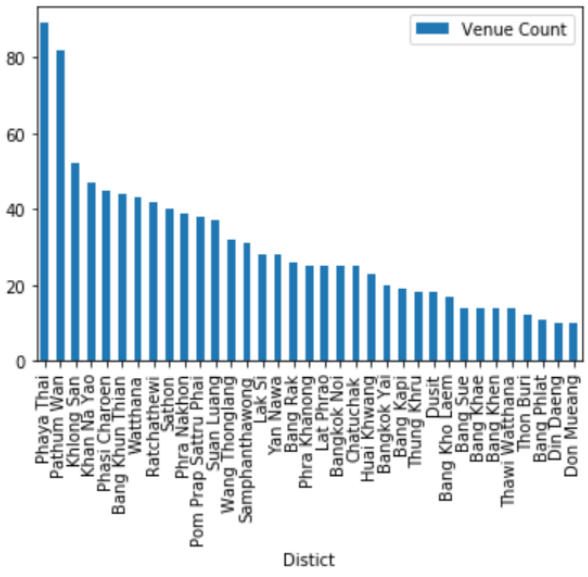
2.3 Feature selection

After data cleaning, there were 50 districts with 164 venue categories. I have data on the population and area size of each district. Then, I calculate population density myself.

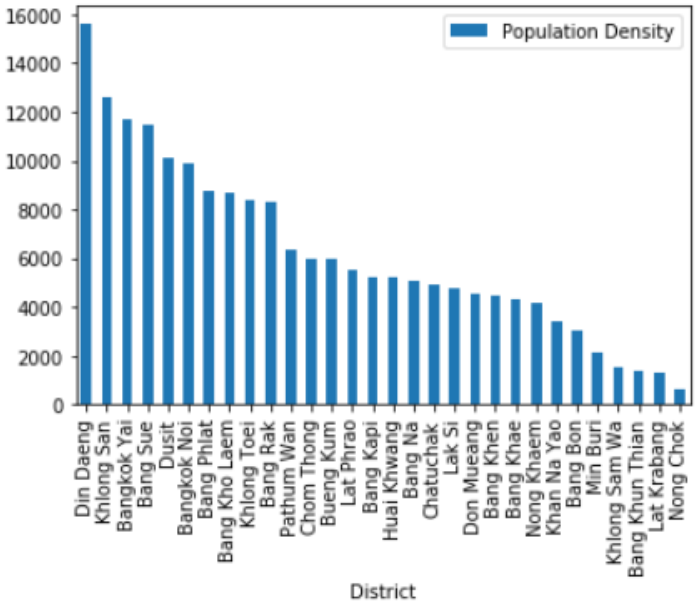
```
Venue Category
Noodle House          126
Convenience Store      80
Coffee Shop            79
Thai Restaurant        74
Café                   53
                      ...
Pastry Shop             1
Gaming Cafe             1
Garden                  1
Garden Center           1
Women's Store           1
Name: Venue Category, Length: 164, dtype: int64
```

3. Exploratory Data Analysis

3.1 Number of venue category count in each district



3.2 Population density in each district

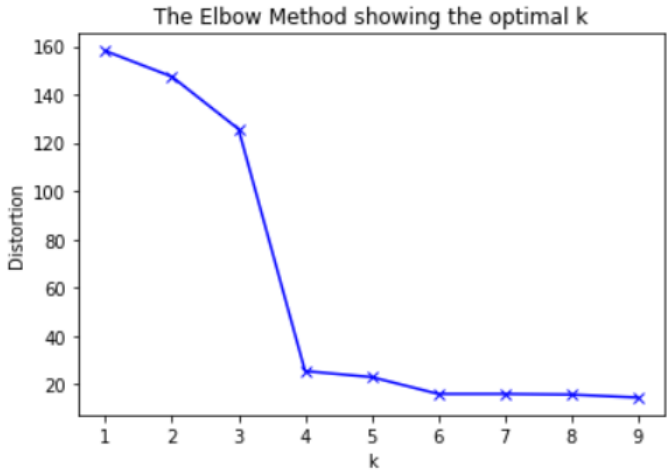


3.3 Sample of top 10 venues in each district

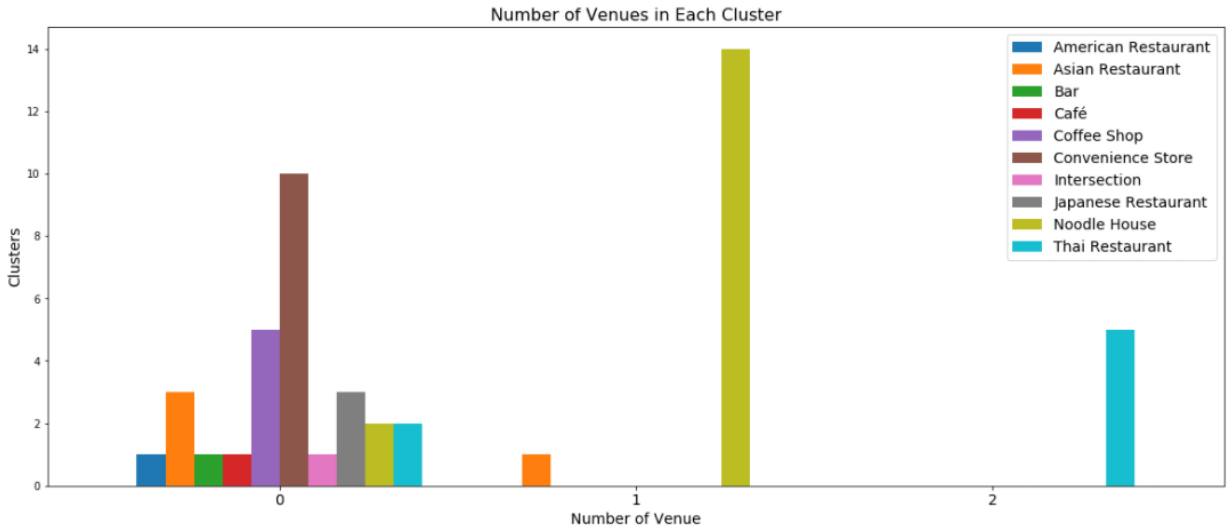
	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bang Bon	Thai Restaurant	Japanese Restaurant	Automotive Shop	Shopping Mall	Convenience Store	Noodle House	Grocery Store	American Restaurant	Optical Shop	Organic Grocery
1	Bang Kapi	Noodle House	Convenience Store	Flea Market	Som Tum Restaurant	Neighborhood	Shabu-Shabu Restaurant	Museum	Multiplex	Coffee Shop	Thai Restaurant
2	Bang Khae	Japanese Restaurant	Convenience Store	Shopping Mall	Noodle House	Asian Restaurant	Pizza Place	Supermarket	Fast Food Restaurant	Coffee Shop	BBQ Joint
3	Bang Khen	Noodle House	Asian Restaurant	Som Tum Restaurant	Convenience Store	Thai Restaurant	Vietnamese Restaurant	Garden	Café	Coffee Shop	Garden Center
4	Bang Kho Laem	Noodle House	Thai Restaurant	Coffee Shop	Chinese Restaurant	Hotpot Restaurant	Vietnamese Restaurant	Shopping Mall	Museum	Fast Food Restaurant	Convenience Store

4. Predictive Modeling

I use K-Means method for create model to clustering. I try to calculate each number of K in iteration to find the best number of K to create model.



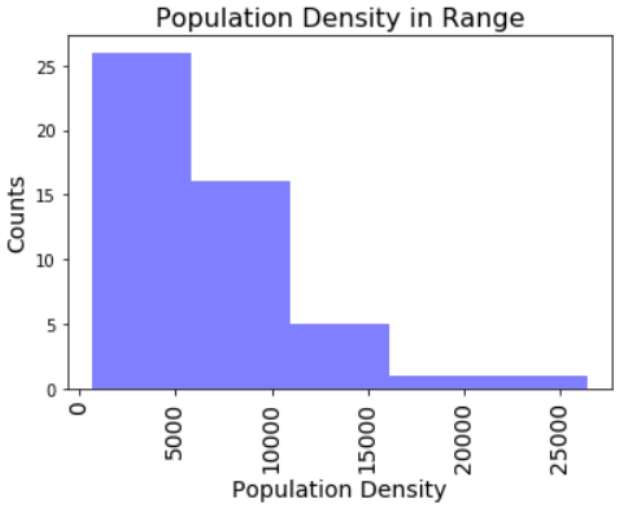
Then, I estimate the number of 1st Most Common Venue in each cluster and also create the chart.



After I examine each cluster, I can label as follows:

	Clusters	Labels
0	0	Multiple Social Venues
1	1	Noodle House Venues
2	2	Thai Restaurant

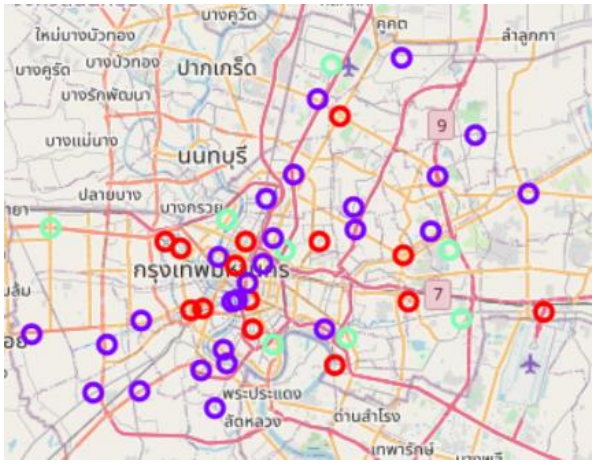
Then, I examine the frequency of population-density by using histogram for visualize.



As it seems in above histogram, I can define the ranges as below:

- “Low Level Population Density” for population density less than 5,000 people per kilometer square
- “Mid-1 Level Population Density” for population density between 5,000 to 10,000 people per kilometer square
- “Mid-2 Level Population Density” for population density between 10,000 to 15,000 people per kilometer square
- “High-1 Level Population Density” for population density between 15,000 to 20,000 people per kilometer square
- “High-2 Level Population Density” for population density more than 20,000 people per kilometer

Map with results



5. Conclusions

In this study, I develop a clustering model for understanding the cluster of each district in Bangkok according to population density and social place density. Although, there are many factors if you are on decision what location you should buy or lend house or condo in Bangkok, for example, your work location, budget, etc. But this model can be one of very useful for support decision.

6. Future directions

However, this model has not enough data for precise clustering. Now, it's only an approximate cluster. But this model is a good basis for improvement and adding more features to more precision in the future.