**Investment in opening a restaurant by population**

**1.Introduction**

**Background**

Currently, there are many food investors. That intends to open a restaurant in Thailand, Therefore there is a problem in choosing the store opening area. Whether it is taste, environment, including wanting to open a shop in the area That have the same or similar restaurant. So presenting the area That has as many foreign restaurants as I can imagine, It is Khlong Toei district.

**Problem**

Choosing to establish a store in the desired area according to the type of shop or suitable location whether competitors or the population in that area is large enough to be noticed or interested in entering the store or not.

**Interest**

Intend to allocate useful information for decision making in establishing a store There are a population of residents and a number of international restaurants in Thailand, including Thai restaurants.

**2. Data acquisition and cleaning**

**2.1 Data sources**

The number of residents in the area is part of the relatively important information. That dataset [here](https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok) , Including image maps from the Foursquare venue categories website [here](https://developer.foursquare.com/docs/resources/categories)

**2.2 Data cleaning**

Data downloaded or scraped from multiple sources were combined into one table.

First, starting with the data from the population of that particular area to present for comparison in making a decision.

Second, select map data for that area only from Foursquare website and select all restaurant from categories, the amount that can be determined according to the website's rights. In which the website has a variety of categories to choose from, this time choosing a restaurant. The problem is that the data of some areas are NaN. It is necessary to change that value to Bangkok or most areas by replacing that data with pandas. After that, the map was simulated. It turns out that that data has come out 50 points from that area. Which I intend to divide the area data by rate so that I know how many people can go to the high-priced stores. But was unable to do because my per-day privileges had already been used up.

My intention is to know the population of that area in the area of ​​the store or the group that has the most. That means that there is a chance that people will live at that restaurant or nearby. Then I chose the K-means Cluster feature to divide the store group. The result is ​​5 neighborhoods and 29 stores type.

**2.3 Feature selection**

After data cleaning, There were 50 samples around that district. Then group rows by neighborhood and by taking the mean of the frequency of occurrence of each category and print each neighborhood along with the most common venues



Figure 1. All data table



Figure 2. All data table after cleaning

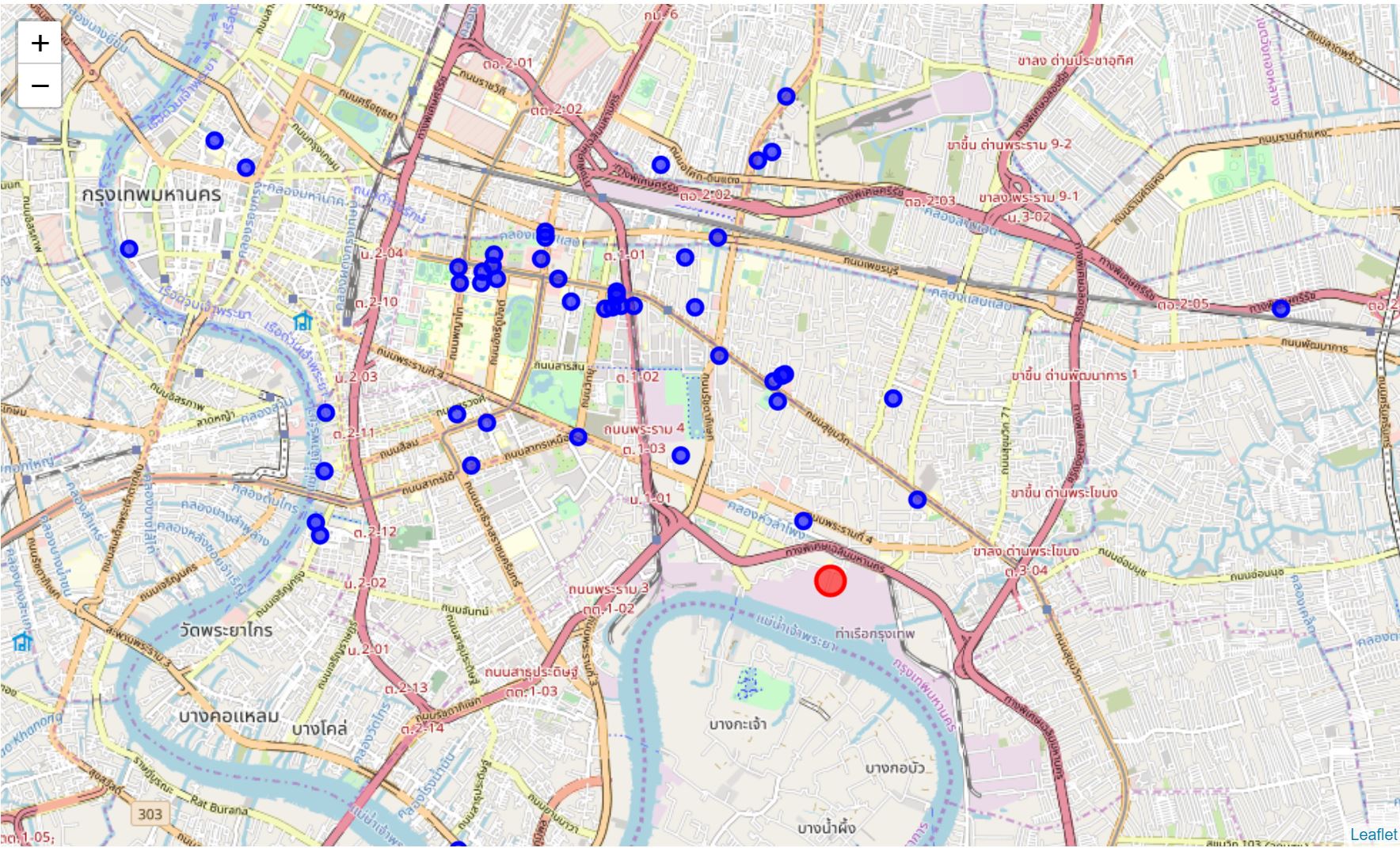


Figure 2. Map

**3. Exploratory Data Analysis**

**3.1 Calculation of target variable**

that area in the area of ​​the store or the group that has the most. That means that there is a chance that people will live at that restaurant or nearby. Then I chose the K-means Cluster feature to divide the store group. The result is ​​6 neighborhoods and 29 stores type.

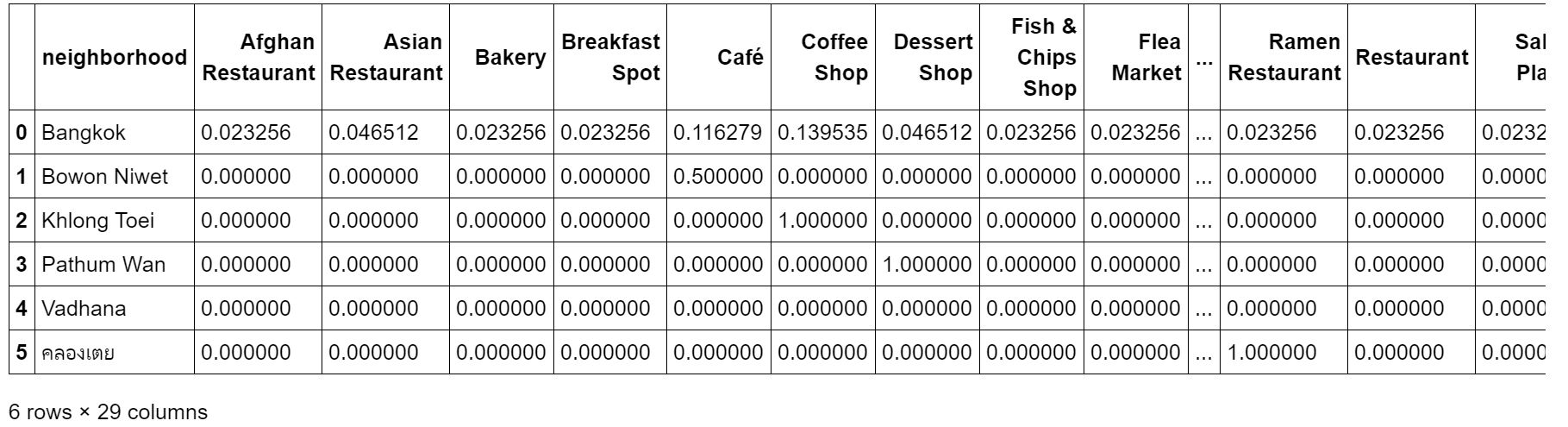


Figure 3. Group rows by neighborhood and by taking the mean of the frequency of occurrence of each category

**4. Predictive Modeling**

These data are stable and therefore choose to use unsupervised calculations. Just group the data of the area.

**4.1 K-means Cluster**

First, Create the new dataframe and display the top 10 venues for each neighborhood.



Figure 4. The Top 10 venues for each neighborhood

Run k-means to cluster the neighborhood into 5 clusters.

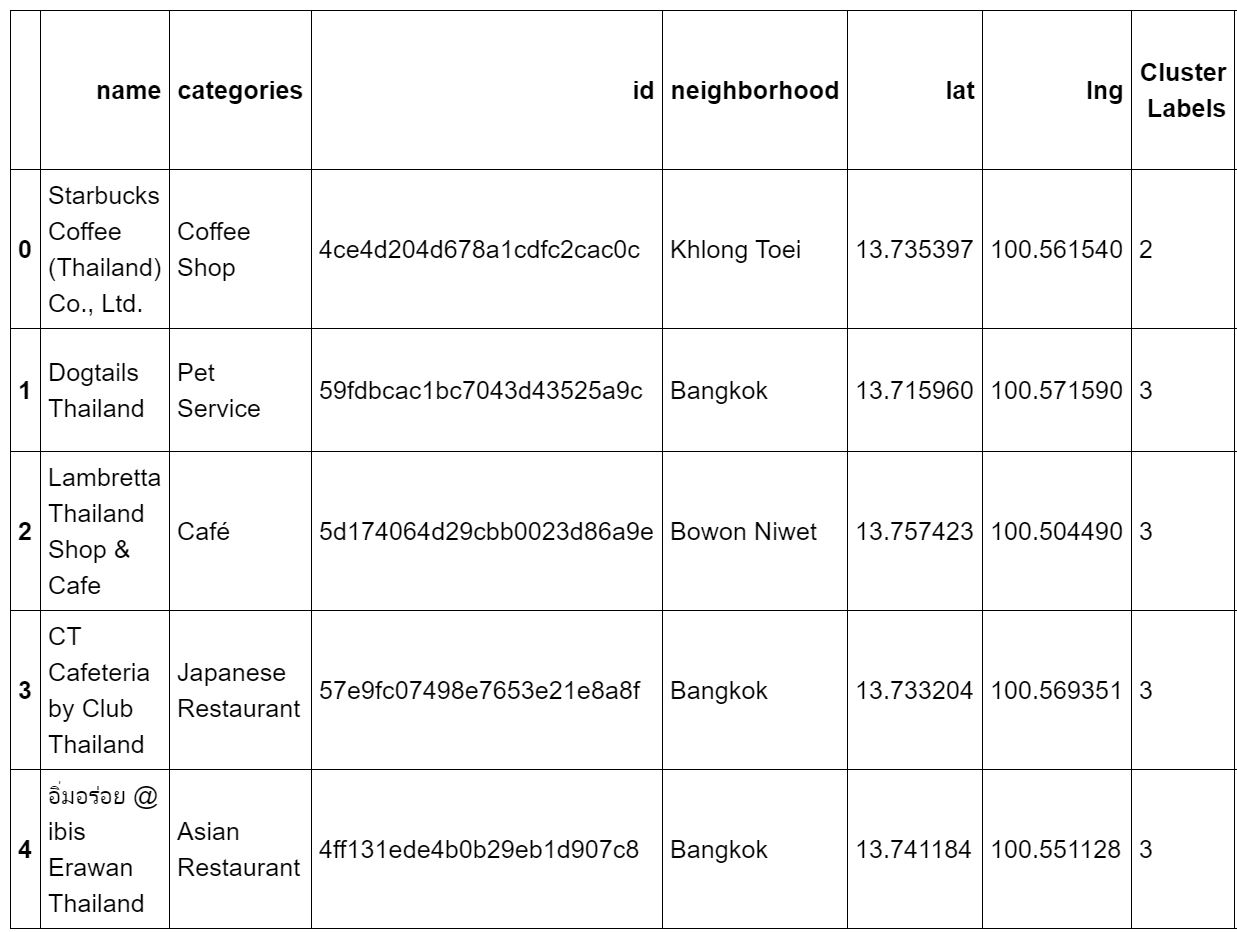


Figure 4. New dataframe that includes the cluster

**5. Conclusions**

In this study, I have analyzed the groups of nearby restaurants according to each neighborhood, which can be divided to see which stores are densely or the most suitable for investment in those areas and can predict that the area As mentioned, there are still a lot of people as well.

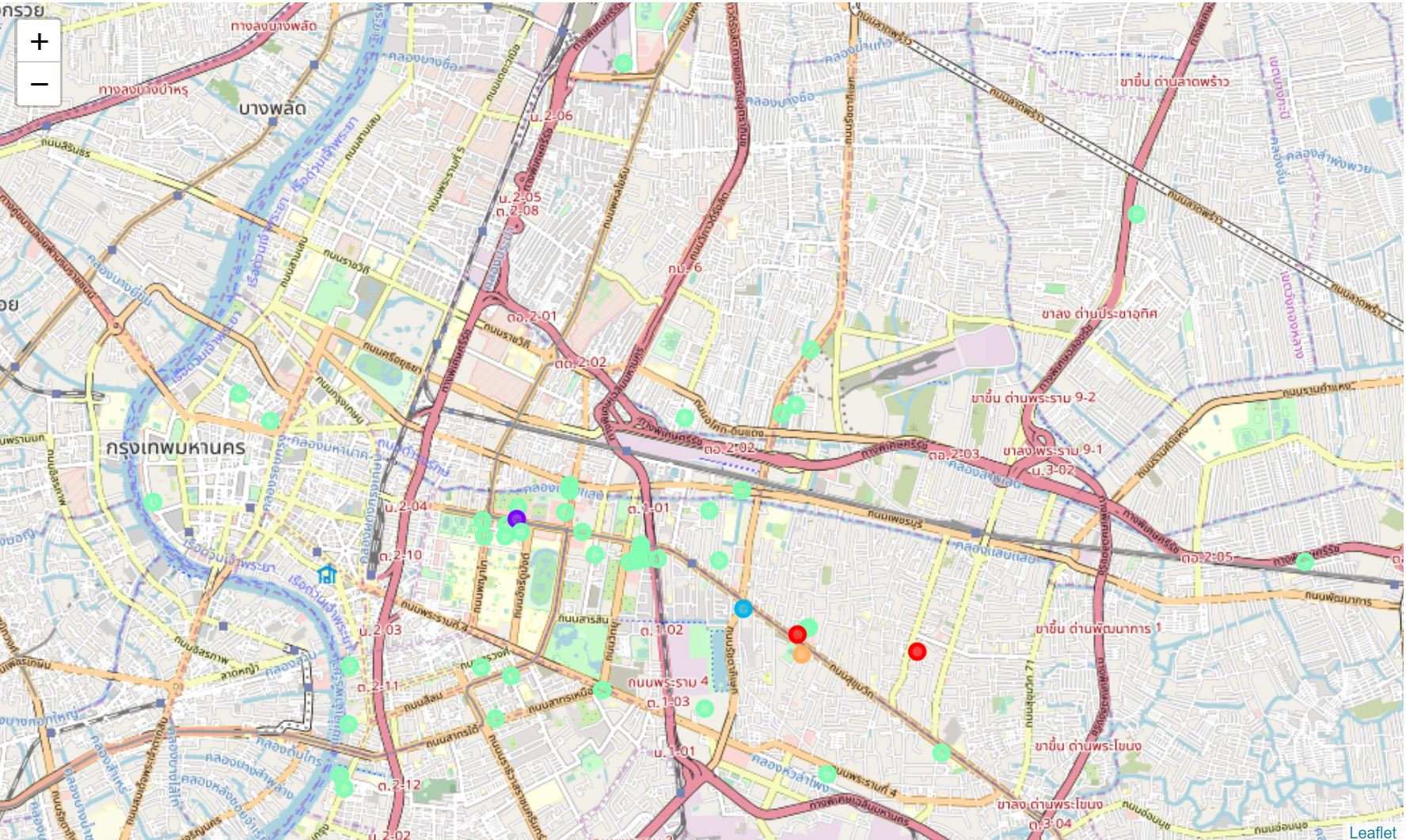


Figure 4. Map after clustering

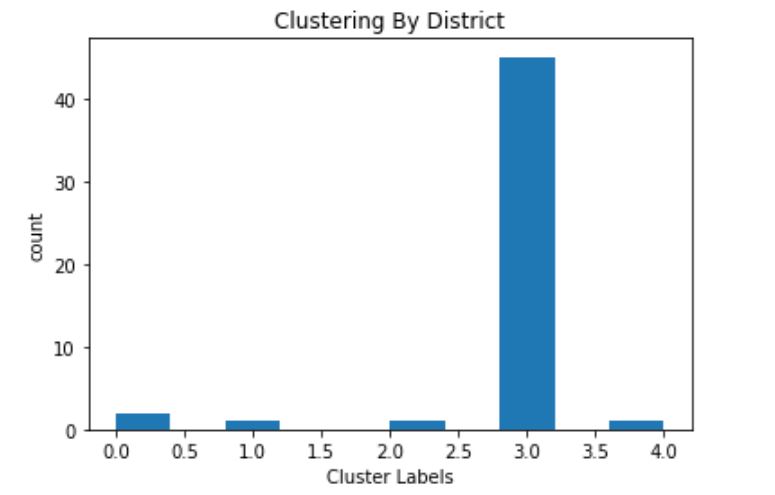


Figure 4. Number of stores in each neighborhood

**6. Future directions**

I see that I intend to develop in the interest rate of each store in order to find out whether the area is suitable for investment in opening a store where there are competitors.