Identifying the best places to open a French restaurant in Nagoya

Jonhathan Pinon

June 30, 2019

1. Introduction

1.1 Background

Nagoya is one of the largest cities in Japan. It has a population of over 2 million and is situated in the middle eastern part of Japan. In Japan, French cuisine is relatively popular. According to Foursquare data, there are about a hundred of French restaurants in Nagoya. Supposing that a chain of restaurants specialized in French cooking want to open new French restaurants in Nagoya, where would be the best places?

1.2 Problem

The best place is the place that would attract the most customers. In this report, we define popularity as the number of customers. Thus, we first need to identify, and ideally quantify, which factors are beneficial for attracting customers, and which are not. For example, we can expect the proximity of subway station to substantially increase the popularity. On the other hand, we can expect that similar restaurants would decrease the popularity. Once those factors are identified and quantified, we will draw a map of Nagoya showing the potential popularity index calculated from the above factors.

1.3 Interest

Such a map would obviously be very interesting for any group who wishes to open a restaurant. A restaurant popularity depends on its quality, but the location is still a major factor. To maximize potential income, choosing the place that would potentially attract the most customers is very important.

2. Data acquisition and cleaning

2.1 Data sources

Foursquare location data will be the only data source. Foursquare provides the position of most restaurants, train stations, subway stations, and other venues in Nagoya. In addition, Foursquare provides stats such the number of current people at a specified venue. Thus, there is no need to use another data source.

2.2 Methodology and features selection

Our study has two parts: first we need to quantify the factors which tend to increase restaurant popularity. Then we need to apply those factors to compute the map of Nagoya showing the potential popularity (just like a temperature map but with potential popularity instead of temperature).

Regarding the first part, we can qualitatively estimate what tend to increase or decrease the popularity of a restaurant. As cited above, we can suppose that the proximity of public transports increase popularity. On the other hand, we can expect that similar restaurants nearby decreases the popularity. However, to provide a better map, we need to quantify the impact of those factors. To do that, we get the list of all French restaurants in Nagoya, and their details. For each French restaurant, we get the list of nearby venues, especially other restaurants and public transport venues.

Unfortunately, the number of people who have been to a specific venue is no longer [available using Foursquare data](). It makes it very difficult to estimate popularity. There is still a way: we can use the *hereNow* field. This field gives the number of people currently at the restaurant. By gathering data over the course of several days, and different hours, we can get a good estimatation of the popularity.

After we know the popularity of each French restaurant, we calculate the impact of the proximity of each venue category on the popularity. In other words, we run a mutiple linear regression with:

- The popularity as the dependant variable
- For the independent variables…:
    - The proximity of other French restaurants
    - The proximity of other European restaurants
    - The proximity of other restaurants
    - The proximity of train stations
    - The proximity of subway stations
    - And so on for every category available using Foursqure data

W propose to quantify the proximity by the sum of the inverse of the distances between each venue and each French restaurant, within one kilometer radius. For example, let us say there is French restaurant named FRA. Let us say there are two restaurants named RESA and RESB. RESA is 400 meters away from FRA. RESB is 800 meters away from FRA. Then the proximity of other restaurants variable is $1/400 + 1/800 = 0.00375$. This way of calculing the 'proximity' may be revised as our model is refined.

Once we have all the coefficients (the impacts of each independent variable on the popularity), we can calculate the potential popularity map for French restaurants in Nagoya. We divide Nagoya in a 1000*1000 grid, and for each sector, we get the nearby venues. Then we apply our model to calculate the popularity of each sector. This is the second and last part of our study.

2.3 Data cleaning

For our study, we just need the list of venues in Nagoya with their category and position. We also need the number of visits for each French restaurant. There are two dataframes. One dataframe for the list of all venues in Nagoya with their category and position. Another dataframe for the French restaurants: each row has a French restaurant with its popularity and location (latitude, longitude), and the 'proximity' of other venues.

We do not expect any missing field. Foursquare always provides the position, category and the number of current users. Foursquare may not list all venues in Nagoya, but this is not a problem. We limit our study to all venues listed by Foursquare, which should be more than enough.