Predicting Best Venue for Opening a Restaurant

Sangyoon Oh

June 26, 2020

I. Introduction

1.1 Background

In South Korea, entertainment flows around the country as a series of trendlines that the public follows religiously. One of the most common aspects of entertainment is food. Koreans follow food from a variety of sources including television, social media, and word of mouth. Whichever dish or visual representation is trending is what the general public veers towards resulting in high revenue for the owners. Being able to predict the next trend or having a consistently high selling product is pertinent to having a successful restaurant.

1.2 Problem

For a new business owner to decide what to sell and where to sell it, it is important to consider a wide variety of factors. The trend of the time, popularity of the cuisine, and pertinence of location. The main feature I will focus on will be location. I will try to pinpoint the best areas on how one can generate the most amount of revenue. I will use data from districts in Seoul, the capital and the largest city in South Korea.

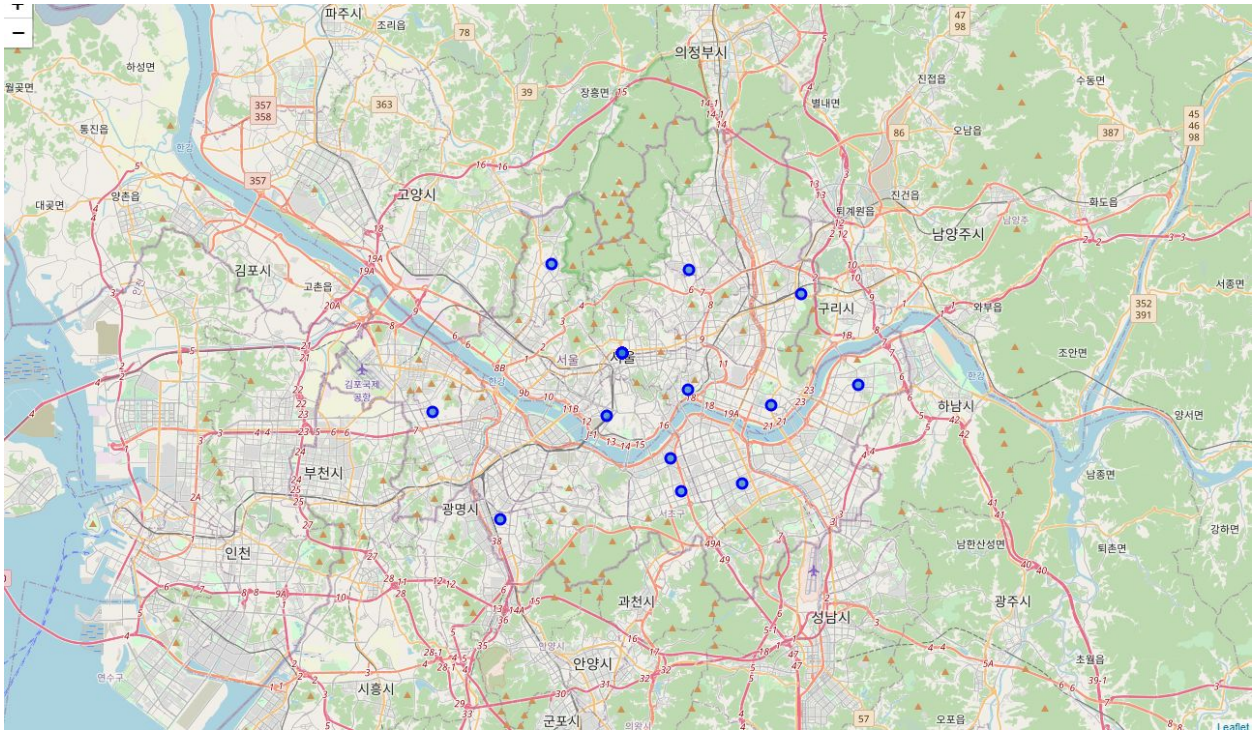
II. Data Acquisition and Cleaning

In order to grasp a general idea of where the most popular restaurants are located, I will locate the geographical locations of each district in Seoul then find the top 50 venues from each district using Foursquare's API. Then I will highlight venues that feature the most popular type of cuisine "Korean" within each district. Finally, using K-Nearest Neighbors, I will set the districts as five separate clusters then check which area is the most heavily competitive and which are less.

III. Methodology

3.1 Initial Mapping

I initially gathered data from a Wikipedia article detailing all the geographical locations of each of the 25 districts of Seoul.



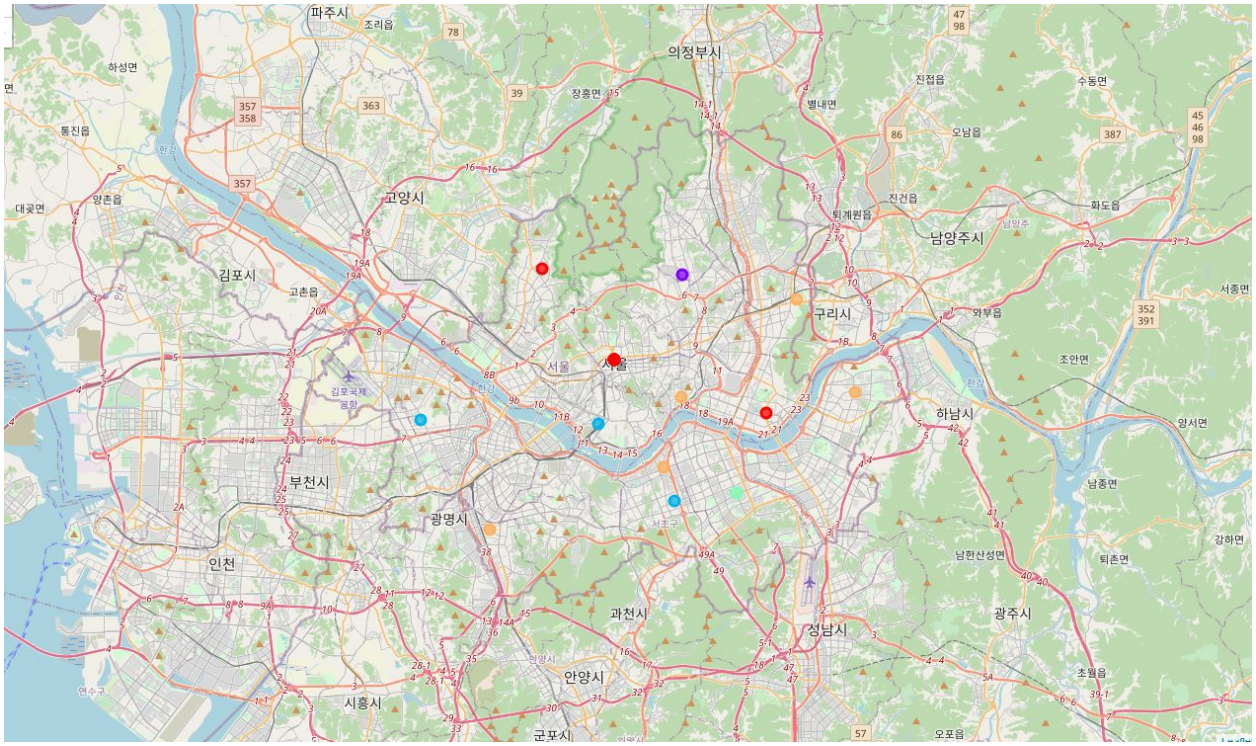
After gathering the longitude and latitude points of all the restaurants, I mapped them out using the Foursquare API. Next using the Foursquare API, I gathered the top 50 venues within a 2000 meter radius of each of the districts. Then I gathered the unique categories of all of the venues that were accumulated.

3.2 Category Organization

Next, using one-hot encoding, I created a dataframe detailing all of the districts with a summary of the categories of the venues. Then choosing the most popular type of restaurant in the dataframe (“Korean Restaurant”), I chose only the column with this cuisine.

3.3 K-Nearest Neighbor

Finally, using K-Nearest Neighbors of 5 clusters, I added the individual cluster numbers to the individual districts. Then each of the clusters were printed out with the frequency of “Korean Restaurant” categories printed out.



IV. Results

The first cluster contained the most number of Korean restaurants and was located near the central area of Seoul as well as the northwestern region. However, cluster number 1 & 3, contained the least number of Korean restaurants. Considering their location; cluster 1 located right on the northern edge of the city and cluster 3 in the southernmost part it is expected it would be the least popular area for restaurants.

V. Discussion

For a new business owner hoping to open up a new restaurant in Seoul, there are benefits and risks to each of the districts. While the most popular cluster (0), seems to be the most populous, the source of competition would be the most severe. A new business opening in those districts might not be able to survive for long due to already established franchises and businesses. On the other hand, opening up a new place in clusters 1 & 3, would avoid the competition, it is also risky due to the fact that not a lot of people tend to go to these districts. While a new restaurant may be able to entice more customers to the area, there are no guarantees that people might flock from the originally established areas.

VI. Conclusion

In conclusion, the city of Seoul is the most populous city in the country. While analyzing the area in which opening up a new restaurant in the city, there were some drawbacks to this analysis. Since Foursquare API is not as widely used in South Korea, there were possible venues and categories which were not included that may have affected the categorization. Ultimately, without running a trial run of the restaurant it is impossible to calculate how successful a new restaurant will be. However, through this data analysis, I was able to provide some background information into where a new owner should try.