Battle of the Neighborhoods

Applied Data Science Capstone Project

Dohun Na

I. Introduction

Lee is an incoming graduate student slated to study at Columbia University in New York. After securing his visa and plane tickets, he now has to decide where to rent an apartment for the duration of his stay. Fortunately, he is not constrained by price, as he and his family are well-off; however, he predicts that during his stay in New York, he will likely be unable to eat non-Korean food in a regular manner. That is, it is essential that he lives in a neighborhood with a) a number of Korean restaurants and b) Korean restaurants that have been given high ratings. Since he will be studying in Upper Manhattan, renting an apartment too far from Manhattan (Queens or Staten Island) is out of the question for Lee. The Bronx and Brooklyn (in addition to Manhattan) are options for Lee, since the 1, 2, and 3 MTA lines all cross these tree boroughs. In an attempt to inform his decision-making, Lee asks John to assist is looking for a neighborhood fitting Lee's requirements. Furthermore, he dreams of opening up his own Korean restaurant as a side gig; thus, he would like to know which borough among Manhattan, Bronx and Brooklyn has the least number of Korean restaurants.

II. Business Questions

Based on Lee’s requirements, the project seeks to answer the following business questions:

1) Which neighborhood in Manhattan, the Bronx and Brooklyn has the highest concentration of Korean restaurants?

2) Which neighborhood in Manhattan, the Bronx and Brooklyn has has Korean restaurants with the highest average ratings?

3) Which neighborhood has the least number of Korean restaurants?

III. Target Audience

This exercise, as well as the outcomes, may be interesting for Korean students who plan to move to the New York, especially those who have a strong affinity for Korean food. Furthermore, the project serves as a rough market study of Korean restaurants for the boroughs of Manhattan, the Bronx and Brooklyn, which could

be of particular interest for restauranteurs interested in opening Korean restaurants of their own.

IV. Data

The following data sources were used for this project:

1) New York neighborhood data to be collected from https://cocl.us/new_york_dataset, and filtered to use only Manhattan, Brooklyn and Bronx areas
2) The number, location and ratings of Korean restaurants will be from FourSquare. After filtering for only Korean restaurants,

The two subsets of data were merged to form one master dataset, of which served as the basis of ranking neighborhood based on Lee's requirements, as well as to visualize the resulting output for maximum impact

V. Methodology

1) Clean and filter New York neighborhood data to reflect only areas of the three boroughs
2) Collect FourSquare data on ratings, tips and likes by users, then filter to show only Korean restaurants
3) Merge the two datasets and sort rankings by borough and neighborhood
4) Visualize rankings

VI. Findings

1. According to Figure 1, we can see that the Bronx has the least number of Korean restaurants, while Manhattan has the most number of Korean restaurants

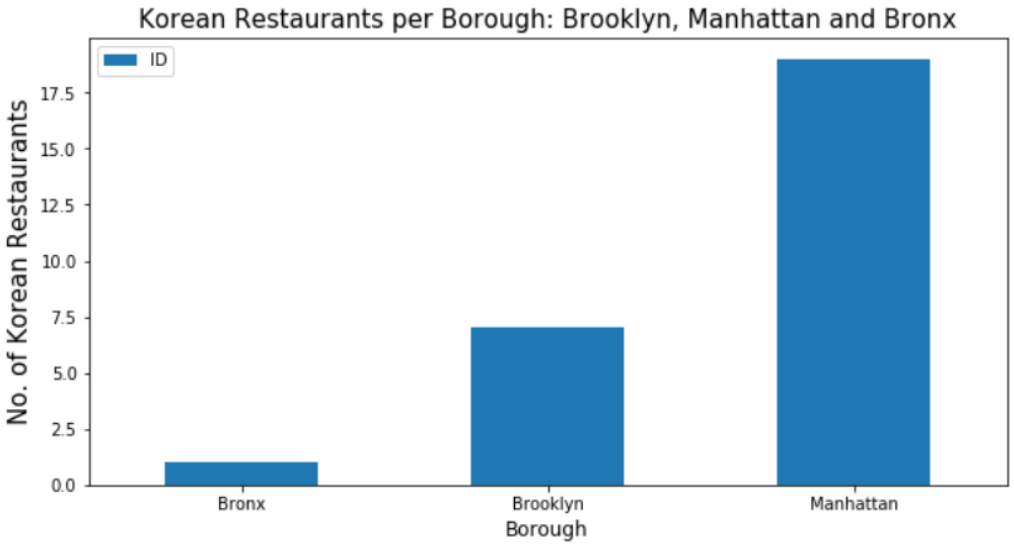

Figure 1

2. According to Figure 2, we observe that on a neighborhood basis, Midtown South has the most number of Korean restaurants. This is expected since the neighborhood is home to Manhattan's K-town

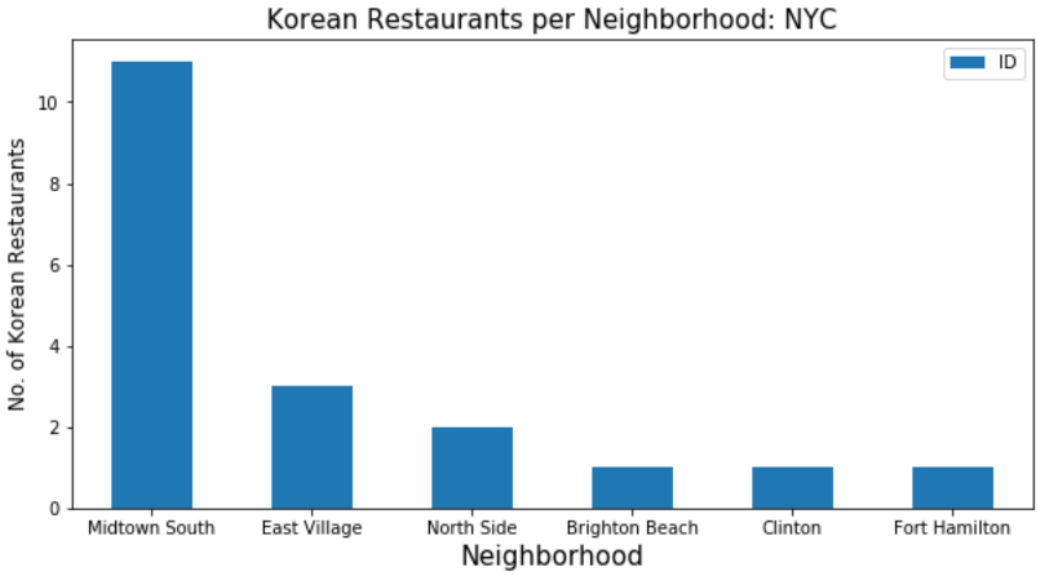

Figure 2

3. Similarly, from Figure 3 we observe that Manhattan has the highest average ratings out of the three boroughs of interest, though Brooklyn is not far behind. The missing data for the Bronx is likely due to unavailable data in FourSquare

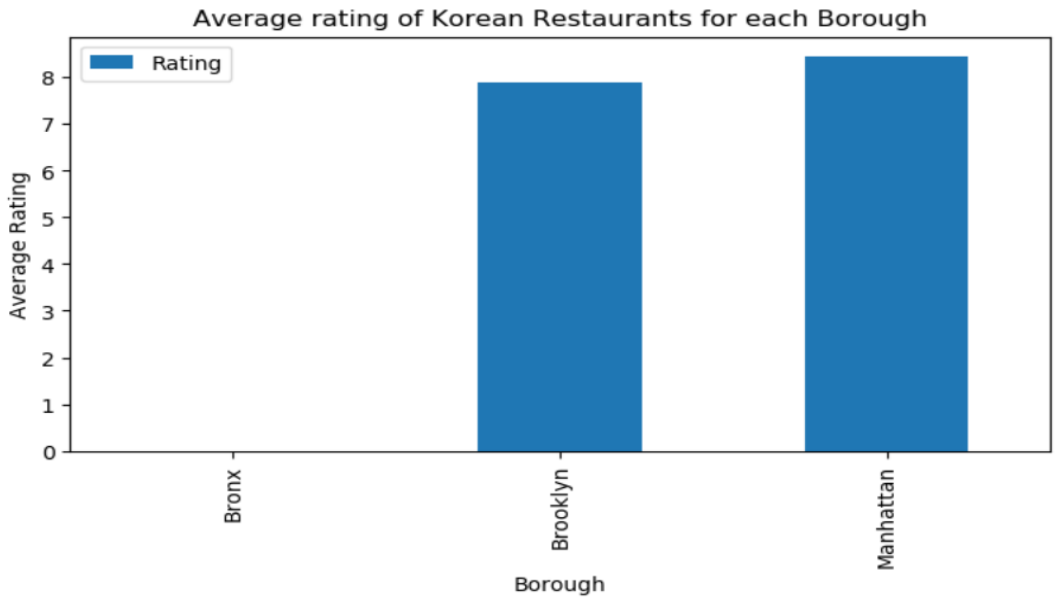

Figure 3

4. However, as seen in Figure 4, the neighborhood with the highest average ratings is not located in Midtown South (the location of Manhattan's K-town), but in the Lower Manhattan neighborhoods of Tribeca and East Village

	Neighborhood	Average Rating
10	Tribeca	8.900000
2	East Village	8.800000
8	Prospect Heights	8.500000
6	Midtown South	8.463636
9	South Side	8.400000
13	Williamsburg	8.400000
1	Clinton	8.300000
12	West Village	8.300000
5	Manhattan Valley	8.100000
7	North Side	8.100000

Figure 4

VII. Discussion

Based on the findings, if Lee decides that having a variety of Korean restaurants within an immediate vicinity is of upmost importance, then searching for an apartment in Midtown South would be recommended. On the other hand, if Lee decides that highly-rated restaurants are his priority, then Tribeca, East Village or Prospect Heights would be recommended. Finally, with respect to his dream of opening up his own restaurant within the three boroughs, it is recommended that he opens one in the Bronx, owing to the relative dearth of Korean restaurants in the general area.

VIII. Limitations

Lee's requirements specifically constrained the areas of interest (down to three boroughs instead of all five); thus, expanding the search to all five boroughs (particularly, the immigrant-heavy area of Queens) may possibly open new windows to more and higher-rated Korean restaurants, although moving to these areas may come at the expense of longer commute times to school.

Furthermore, a free account was used to gather data from FourSquare, hence limiting the amount of API calls, as well as the retrievable data. This means that menu items, pictures of restaurants and food, among others, were not accounted for in the analysis. Future comprehensive research in terms of aiding Korean students will require the use of paid-services in FourSquare.

IX. Conclusion

The dominance of Midtown South (K-town) was apparent in the number of Korean restaurants per neighborhood. However, the exercise unexpectedly showed that, according to FourSquare users, the highest rated restaurants were not in fact in Midtown South, but in Lower Manhattan. Overall, our analysis has provided Lee (and other Koreans contemplating where to rent apartments in NYC) abundant data in this respect, which will inform his decision-making.