The Restaurant Profile of Bronx Area

1. Introduction

1.1 Introduction

Each city may have thousands restaurants of different types. The restaurants usually not distributed evenly in neighborhoods, which add extra complexity to the profile of restaurant. Having a restaurant profile that describes the distribution of each type of restaurant will provide guidance for people to start new restaurants. In addition, the profile of restaurant will also be helpful for people who plan to move to Bronx to choose the right neighborhood based on their food preference. Here, I did analysis for Bronx area as a proof of principle.

1.2 Goal

The goal of this project is to find out which types of restaurants are most popular in Bronx area and which neighborhood is lacking those types of restaurants.

1.3 Interest

- People who want start a new restaurant but not sure which is the best place to start.
- People who plan to move to a city and have strong food preferences.

2. Data acquisition

Two parts of data will be used for this analysis: neighborhood data and restaurant data. The neighborhood data of Bronx are available at https://cocl.us/new_york_dataset. The restaurant data is available through Foursquare API (https://developer.foursquare.com). Foursquare will provide the geological location of restaurants and their the categories.

3. Methodology

3.1 data cleaning and exploratory data analysis

The Jason file that describes New York neighborhood was downloaded and parsed to obtain longitude and latitude data for each neighborhood. In total there are 52 neighborhoods in Bronx area. With the longitude and latitude data, I requested venue data from Foursquare.

I extracted all restaurants from venue data and counted each type of restaurant. I ranked all types of restaurants based on the number and acquired the top 10 most popular restaurant categories (Figure 1). The top 10 restaurant categories account for 78% of all restaurants sampled in Bronx area.

Focusing on the top 10 restaurant categories, I calculated the percentage of each restaurant category in all 52 neighborhoods. I substracted the percentage with the overall percentage in Bronx area to get an overrepresentation value.

Since there are as many as 52 neighborhoods that are showing complex patterns, I applied an unsupervised learn method, hierarchical clustering, to group them based on their overrepresentation value. In this way, it is easier to interpret the data.

Figure 1
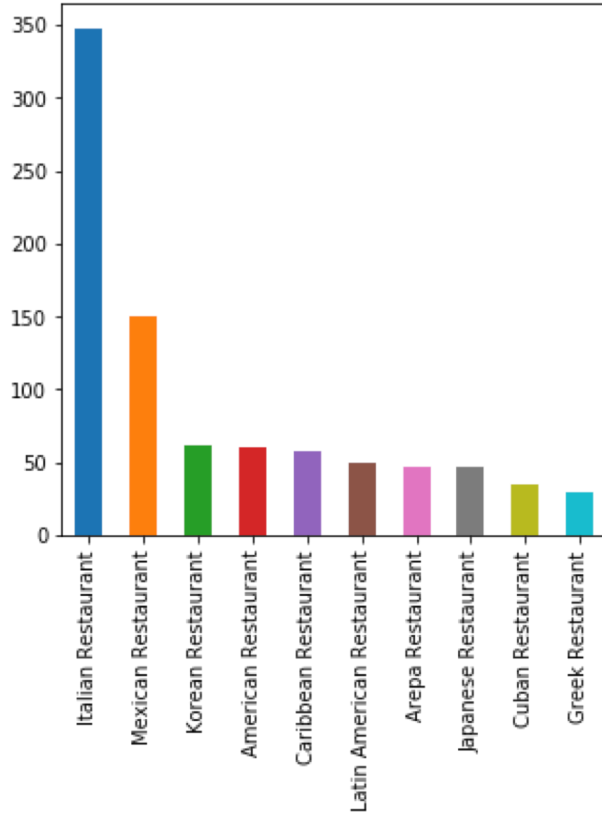
4. Results

The overrepresentation values for each neighborhood are shown as a heatmap (Figure2). Red color indicates the restaurant category is overrepresented in that neighborhood and may have intensive competition. Blue color indicates the restaurant category is underrepresented and may have opportunity to start a new restaurant. The dendrogram indicates the result of hierarchical clustering. The neighborhoods fall into three major clusters. The left cluster contains neighborhoods like Western Square, which are saturated for Italian restaurants and Caribbean restaurants. The middle cluster contains neighborhoods like High Bridge, which are saturated for Korean restaurants. The right cluster contains neighborhoods like Clason Point, which are relatively less Italian restaurants and Mexican restaurants than the average.

Figure 2
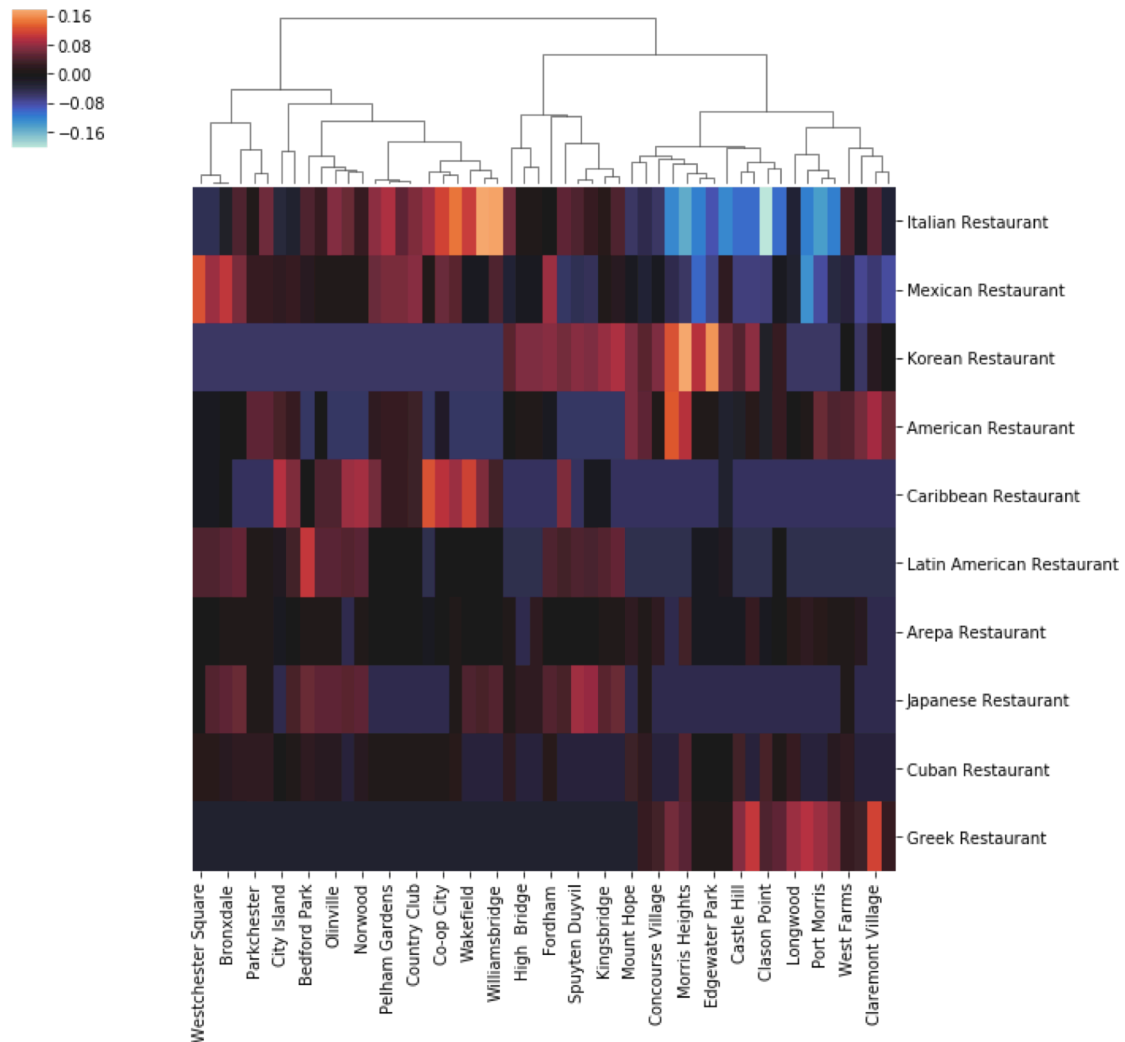
5. Discussion

Based on the hierarchical clustering analysis, the neighborhoods primarily fall into three groups. For the first two groups, some restaurant categories are overrepresented, indicating strong competition. People who want to start a restaurant in those neighborhoods should avoid certain restaurant categories such as Italian restaurant, Caribbean restaurant, Korean restaurant. For the third group, the Italian restaurants and Mexican restaurants are underrepresented. This indicates business opportunities in those neighborhoods.

6. Conclusion

In this analysis, I studies the restaurant profile of Bronx area and provided guidance for people who want to start a new restaurant. We discussed which restaurant categories are recommended/ not recommended for each neighborhoods. The method of analysis can be adapted to analyze any other cities.