Restaurant Clusters

Martin Monsch

March 27, 2019

1. Background and Problem

1.1. Background

Restaurants and bars are usually not evenly distributed in cities. Typically, there are areas with a lot of restaurants and bars, and both locals and tourists crowd these regions in the evenings. Some of these clusters are well-established, some hip, and others are filled with tourist traps. For individuals — especially for tourists — it is not always evident where the restaurant and bar clusters are and how they qualify. City guides or blogs provide information, but the acquisition of the required information is troublesome and reliable information rare. Even many locals are often not up-to-date and do not know where the good restaurants are located since restaurant clusters that were hip once might have already become well-established or even an area filled with tourist traps.

1.2. Problem

The first goal of the capstone project is, therefore, to find a way to cluster restaurants based on their location to identify and visualize restaurant clusters for a given city. As a second goal, the average rating of restaurants in a specific cluster shall be determined and the different clusters visualized again based on the different average ratings to make it easy to find restaurant clusters with a high (or low) average rating.

1.3. Interested Parties

The restaurant clusters are of interest for the following groups: (a) consumers, (b) venues, (c) investors and investment companies, (d) city planners, and (e) media companies. For consumers, the clusters are especially interesting if they do not know the city well, which is the case for tourists, other travelers, and newly settled individuals, but also locals may be interested to stay up-to-date and to know where new clusters emerge. For venues, such as restaurants, the clusters can help to determine where to locate a restaurant or a bar, whether the goal is to be within a well-rated cluster, a not so well rated-cluster for the sake of less competition, or in an area in between clusters with an undersupply of restaurants. Investors can potentially get relevant information from these clusters as well, especially if they are compared over time to identify trends, and city planners might find the clusters useful for their work. Finally, media companies could be interested and companies that edit tourist guides.

2. Data Description and Use

To build the desired restaurant clusters, location data and ratings for restaurants will be required. This data will be collected through the Foursquare API. New York has been chosen as the location for the clusters, but it should be possible to apply the same algorithms to any other city. Since both the search and explore queries of the Foursquare API are limited to fifty venues, a relatively dense grid with location points will be required to collect sufficient data on restaurants in New York. The collected data will then be merged into one data frame and duplicates eliminated. The ratings will be added at a later

stage and just for a sample of restaurants since my access to rating data is limited to 500 venues per day. The required columns are “ID” (to get the ratings later), “Venue Name”, “Venue Category” (since filters may be applied at a later stage), “Latitude”, and “Longitude”. For the clustering of the restaurants, only the latitude and longitude for all the restaurants are required. As a clustering algorithm, DBSCAN will be used since this algorithm identifies outliers and finds clusters with an uneven shape. The clusters will then be visualized using Folium.

As a next step, a sample of clusters and venues will be selected and used to access rating data through the Foursquare API. Then, the data frame will be grouped by the different clusters and the mean rating calculated for each cluster. This average will then be inserted as a new column in the data frame with the sample of restaurants. For the visualization, the color map RdYlGn will be used to visualize restaurant clusters with good ratings dark-green and clusters with bad ratings dark-red.