Outlining Restaurant Clusters in Beijing, China

Jiachen Wei
20695056

A Senior Honours Thesis Submitted in Partial Fulfillment of the Degree of Bachelor of Environmental Studies (Honours Geomatics)

Department of Geography and Environmental Management
Faculty of Environment
University of Waterloo

March, 2018

Table of Contents

List of Figures

List of Tables

Abstract

Cluster is a significant spatial pattern, and it can help local businesses make geographical decisions. The thesis aims to describe the global point pattern of the restaurants in Beijing and to identify significant local clusters.

The global point pattern is described by visual examination, comparison to population in terms of count and density, centrography, and Moran's I statistic. These methods find a significant cluster at central Beijing and a directional trend towards the northeast.

To identify significant local restaurant clusters, the thesis considers the Getis-Ord G*i* statistic and Anselin local Moran's I for the quadrat count values. As the statistically significant quadrats show clustering pattern in the result, the thesis continues to describe this pattern using standard deviational ellipse. Such an ellipse is defined as the outline of a major restaurant cluster (MRC). Different criteria are applied to select the quadrat members of MRC. Finally, the thesis delineated 28 MRCs and suggests possible applications of these geographic boundaries.

1. Introduction

1.1 Motivation

Online-To-Offline (O2O) restaurants, one of the most representative successes in China's O2O e-commerce, reached a market size of 94.4 billion RMB in 2014 (Liu & Xu, 2016). According to Pan, Wu, & Olson (2017), O2O refers to the business approach that attracts customers to physical stores from online platforms including review, navigation, and payment. Chinese O2O restaurants usually display cuisine information including review and average spending on website or mobile application similar to Yelp.

The sizeable market of O2O restaurant may interest local administrators and prospective investors. O2O platforms present open data in the form of text (for example, restaurant locations and cuisine type), and the location-based data hold potential for spatial analysis to understand the distribution of trade areas and the location strategies. Popular O2O platforms are believed to be a reliable data source because they are frequently used by a large number of people. The thesis considers restaurant information from Dianping, the O2O platform with the largest number of active consumers and merchants in China (Xiao, Mi, Zhang, & Ma, 2017).

Another motivation of the thesis comes from the concept of shangqu. Many Chinese O2O platforms allow users to filter restaurants by shangqu as an approximate location that interests the users. Shangqu literally means a trade area or a business area, which implies that local restaurants are more densely located in some areas than the others. Different from an administrative division or a commercial zoning parcel, the boundary of a shangqu is not defined.

1.2 Research Objectives

The thesis contributes to the understanding of how restaurants are located in Beijing, China. The objectives of the thesis are listed as follows.

A. Describe the global clustering pattern of restaurants in Beijing. The thesis assumes that clustering patterns of restaurants exist in Beijing because popular O2O platforms provides many shangqu filters for local users.

B. Identify significant cluster of restaurants and delineate the clusters of considerable size.

Fulfilment of Objective B will be valuable for different audiences. Administers in the restaurant industry can allocate more human resources for areas with more restaurants, or encourage local businesses in the catering industry. For investors and restaurant owners, identifying the dynamic restaurant locations will be beneficial to their location strategies. They may also be able to study customer preferences within certain geographic areas. In addition, travelers or new workers in Beijing can choose to stay in areas close to the restaurant clusters to take advantage of the commercial convenience.

1.3 Thesis Outline

Section 2 introduces the study area with a focus on the city's division of functional zones.

Section 3 defines the important concepts in the thesis for the reader's convenience.

Section 4 is an overview of relevant academic works. Section 4.1 explains why cluster is an important subject of study, and Section 4.2 reviews the approaches that can be taken to fulfil the research objectives.

Section 5 related the literature review to the thesis in terms of why the objective is important and how it can be achieved.

Section 6 lists the equipment for the study and provides every step of the data collection and data analysis.

Section 7 presents the results and interpret them in response to the research objectives.

Section 8 summarizes the results in terms of how they fulfil the research objectives.

Section 9 reflects on the limitations in the equipment and methods applied by the thesis.

Section 10 suggests possible applications of the thesis findings.

2. Study Area

The study area is China's capital city, Beijing. According to Beijing Municipal Bureau of Statistics (2018), Beijing currently consists of 16 districts, as a result of merging 7 urban districts in 1985 with the 9 peripheral counties. The 16 districts are divided into 4 zones according to their functions (Figure 1), which are summarized in Table 1 from Beijing Statistical Yearbook (Beijing Municipal Bureau of Statistics, 2017).

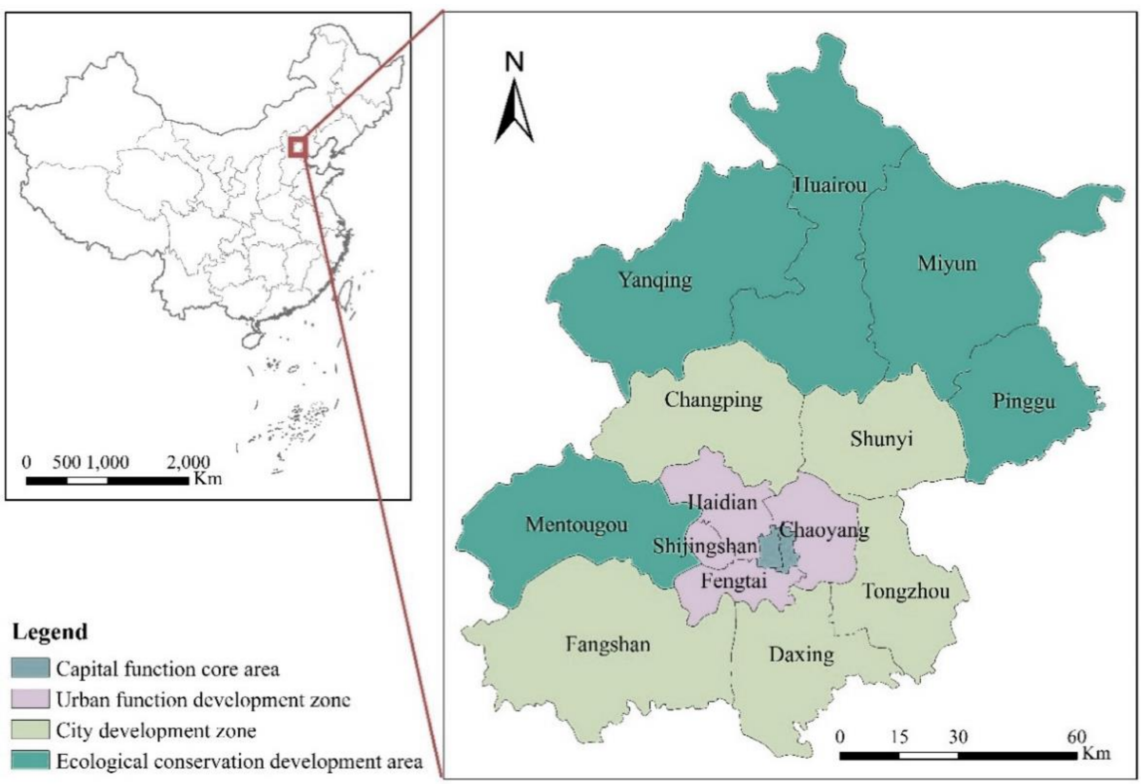


Figure 1 Beijing Administrative Divisions (Qi, Gao, & Zhang, 2017)

Table 1 Definition of Beijing Functional Zones (Beijing Municipal Bureau of Statistics, 2017)

Zone	Districts	Functions
Core Functional Area of the Capital	Dongcheng and Xicheng	National government and historic sites
Urban Function Extension Area	Chaoyang, Fengtai, Shijingshan, and Haidian	Commercial, high-tech industries, and Olympics
New Area of Urban Development	Fangshan, Daxing, Tongzhou, Shunyi, and Changping	Manufacture and agriculture
Ecological Conservation Area	Mentougou, Yanqing, Miyun Huairou, and Pinggu	Mountainous areas for conservation

3. Glossary of Terms and Abbreviations

Dianping: a Chinese equivalent to Yelp. It is an online platform where users explore and review local businesses.

GIS: geographic information system.

Major restaurant cluster (MRC): a cluster of the statistically significant quadrats identified in Section 7.3.1. An MRC is delineated by a standard deviational ellipse of a group of neighboring quadrats of statistical significance. The thesis defines that the boundaries of the ellipses are not intersected.

Online-To-Offline (O2O): the business approach that attracts customers to physical stores from online platforms (Pan et al., 2017).

Point event: the occurrence of a point object. In the thesis, a point event refers to a restaurant.

Point pattern: a spatial pattern that contains only point objects.

Primary filter (PF): a filter on Dianping website to select restaurants by their cuisine origin or food type. See Figure 2.

Secondary filter (SF): after selecting some of the primary filters, users of Dianping are provided with subordinate filters to continue narrowing down their range of choice. See Figure 2.

Shangqu: literally meaning “a trade area” or “a business area” in Chinese. This concept is used by O2O platforms to allow users to filter restaurants by location. The boundary of a shangqu is not defined.

StD: an N (N=1,2,3) standard deviational (NStD) ellipse is centered at the geometric mean, with the length of its semi-major axis equal to N times the standard distance (average of coordinate differences from the mean in the same direction).

Trade area: the geographic area from which a retail store draws most of its customers (Ghosh & MacLafferty, 1987).

4. Literature Review

This section reviews classic works on relevant concepts and recent peer-reviewed journal papers on relevant topics. Section 4.1 provides explanations of trade area (or shangqu in Chinese) and clusters. Section 4.2 focuses on the GIS analysis approaches to identifying restaurant clusters in previous studies in order for the thesis to achieve the research objectives.

4.1 Trade Area and Cluster

Location analysis and location strategy are important for retail businesses including restaurants. As Taneja (1999) famously said, the three critical elements of retail success are "location, location, and location". A good location is crucial for attracting customers and beneficial to enhancing customer loyalty (Craig, 1984; Ramanathan & Ramanathan, 2011).

One component of location analysis is to analyze the trade area, or shangqu in Chinese. Ghosh & MacLafferty (1987) define a trade area as the geographic area from which a retail store draws most of its customers. Wang (2014) believes that trade area analysis not only allows new stores to identify opportunities and competitors, but also helps existing businesses determine their promotional focus and evaluate their performance. However, the thesis is not likely to be able to obtain data on where restaurant customers come from.

Instead of trade areas of restaurants, the thesis considers restaurant clusters. In a commercial context, Porter (2000) defines a cluster as a geographic concentration of interconnected companies that are linked by commonalities and complementarities. Cluster represents an important way of thinking. Well-known examples include John Snow’s discovery of clustering cholera incidents in London and the study of agglomeration in economics (Piore & Sabel, 1984). Porter (2000) associates clusters with competition and cooperation, and competitive advantages in productivity and innovation.

Cluster is a spatial pattern of significance for restaurants. It suggests that high demand exists in an area for the businesses to be profitable (Craig, 1984). This area may also imply the acceptable travelling distance of the customer base. For example, Long & Thill (2015) applied an estimated 750 walking distance from bus stop in their study in Beijing. This static approach is limited in accounting for differences in the cluster size and product offering (i.e. cuisine type for restaurants) across the study area. Moreover, cluster can increase the attractiveness as a “visible restaurant district”, which is one of the benefits from agglomeration effects (Smith, 1985). However, the success of a restaurant is not guaranteed by merely a good location, the product and service it offers matter as well. Litz & Rajaguru (2008) find that small stores offering specific products or services are able to succeed without being located in attractive areas of agglomeration effects.

Cluster is worthier of attention than the other types of point patterns—random and regular. Gould (1970) suggests that point patterns are almost never random as point events tend to show spatial autocorrelation. O'sullivan & Unwin (2014) point out that it is usually more important to detect and locate significant clusters than to test and declare a point pattern as clustered, random, or regular. Hot Spot Analysis (ArcGIS, 2016b) is one tool to identify statistically significant clusters of high values (hot spots) and low values (cold spots).

4.2 Studies on Point Pattern Analysis

This section reviews previous studies on point pattern analysis with a GIS component. It explains the concepts and approaches that are applicable to the thesis and reflects on relevant studies.

4.2.1 Describing Global Point Patterns

O'sullivan & Unwin (2014) point out the distinction between first-order effect and second-order effect for describing a point pattern. First-order effect emphasizes the variation in event intensity across the entire study area, whereas second-order effect is manifest as variation in distance between nearby events. The thesis will briefly examine the first-order effect of the restaurant point pattern in Beijing and focus on cluster detection.

Density describes the first-order properties of a point pattern. Compared to overall density of events which is coarse and subject to the modifiable definition of the study area, exhaustive quadrat count is more commonly used to reveal the spatial variation in density within the study area (O'sullivan & Unwin, 2014). Despite the challenge to set the quadrat size for the entire study area, the thesis can operationalize point clusters as quadrats with significantly high point density.

One description of the second-order properties is whether the overall pattern is clustered. Most commonly methods are based on distance. K-function explores the presence and measures the scale of spatial clusters (Ripley, 1976). Indices based on nearest neighbor provide similar function to K-function while allowing more variation in spatial weights (Fritz, Schuurman, Robertson, & Lear, 2013). Moran’s I statistic measures the global autocorrelation of locations and values together using distance-based spatial weights (O'sullivan & Unwin, 2014).

Centrography is another approach to describe a point pattern. Similar to basic statistics, centrography describes a set of point events using the mean center (mean coordinate values) and the standard distance from the mean center. Plotting a summary circle based on the mean center and the standard distance can visually present the dispersion of events around the mean center. ArcGIS (2016 a) is able to characterize vector features with a standard deviation ellipse, which reveals the directional trend in dispersion rather than applying the same standard distance to all directions.

4.2.2 Identifying Local Clusters

Research Objective B is to outline major local clusters. There two commonly used statistical methods to identify “significant” clusters are the Getis-Ord Gi statistic and Anselin Local Moran’s I.

The G*i* statistic detects local concentrations of high or low values in one attribute of a collection of geographic features. For a location *i*, the G*i* value equals to the weighted sum of the attribute values at all the other locations divided by the sum of the attribute values at all the other locations (Ord & Getis, 1995). One common method to weight the locations is to use the inverse distance from location *i*. The G*i* value is relatively high for a significant cluster of high values (a hot spot), and relatively low in a cold spot. The ArcGIS (2016 b) Hot Spot Analysis tool considers Getis-Ord G*i** statistic, which is defined similarly to G*i* but including location *i* in the calculation. Along with the G*i** statistic, the tool produces confidence levels for each feature to reject the randomization hypothesis.

Local Moran's I indicates four types of local spatial associations. Anselin (1995) modified Moran's I statistic to examine the neighborhood of each geographic feature rather than the entire data set. Based on local Moran's I, the ArcGIS (2017) Cluster and Outlier Analysis tool can identify a significant high-value cluster surrounded by high or low values, or a low-value cluster surrounded by high or low values, by examining positive and negative autocorrelation in the Moran scatterplot. A positive Moran's I indicates the feature belongs to the surrounding cluster while a negative Moran's I are assigned to the outliers. Getis & Ord (1992) encourage the use of both the Gi statistic and the Local Moran's I as the two approches focus on aspects to measuring spatial clusters.

5. Research Rationale

Cluster and trade area are both important for retail businesses. As it is difficult to obtain customer mobility data to identify effective trade areas for the restaurants, the thesis focuses on delineating clusters in the existing restaurants. With approximate boundaries identified for the restaurant clusters, future studies will be able to investigate into topics including retail competition, investment opportunities, land use, customer preferences, and business promotion within each of the areas. These issues are expected to be different across space due to varying population structure and urbanization intensity.

Cluster itself is also an interesting subject of spatial analysis. The thesis will first examine the point pattern over the entire study area in terms of density and autocorrelation. This point pattern is expected to show clusters because most geographic events are spatially autocorrelated and because the O2O platforms suggest clusters by filtering restaurants by shangqu.

With understanding of the global pattern of the restaurants (Research Objective A), the focus of the thesis is to identify and delineate major local clusters (Research Objective B). In light of the literature, common statistical approaches will be taken to identify significant clusters, and a group of neighboring clusters can be characterized by an standard deviational ellipse.

6. Procedure

This section lists the steps in data collection and data analysis. The thesis uses WebHarvy 5.2 to scrape restaurant data, ArcMap 10.5 to process and visualize spatial data, and Excel 2016 to process text data. UTF-8 encoding is applied to all the steps.

6.1 Collecting and Preprocessing Data

The thesis collects restaurant information in text format through web scraping, and downloads public geospatial and statistical data. Next, the restaurant text data were converted into points based on restaurant addresses. The geocoded points will be analyzed with the public data to accomplish the research objectives.

6.1.1 Restaurant Data

The thesis considers Dianping as the optimal data source for restaurant information, for its detailed classification of restaurants and high data quality maintained by active users. Dianping, a Chinese equivalent to Yelp, is an online platform where users explore and review local businesses including food, accommodation, health, and entertainment. After its merger with a similar Chinese service called Meituan in 2015, Dianping/Meituan is estimated to hold an 80% market share of China’s online-to-offline restaurant services (Ma, 2017). Xiao et al. (2017) believe that Dianping has the largest number of active consumers and merchants.

The restaurant information was crawled from Dianping using WebHarvy 5.2 from February 1 to 5 in 2018. Referring to Figure 2, The Dianping website for Beijing allows browsers to view restaurants by category using 36 primary filters (PFs). One of the PFs is “fresh groceries”, and merchant information under this category is not collected for the thesis. Some PFs contain secondary filters (SFs) while others do not (see Appendix A for the list of PFs and SFs considered by the thesis). A total of 34485 entries of data were collected after applying each of the 43 origin-based filters, and each of the 32 food-based filters. The location filters include popular shangqus, administrative divisions, and subway stations. The recommendation filters and location filters were not applied to avoid duplicates.

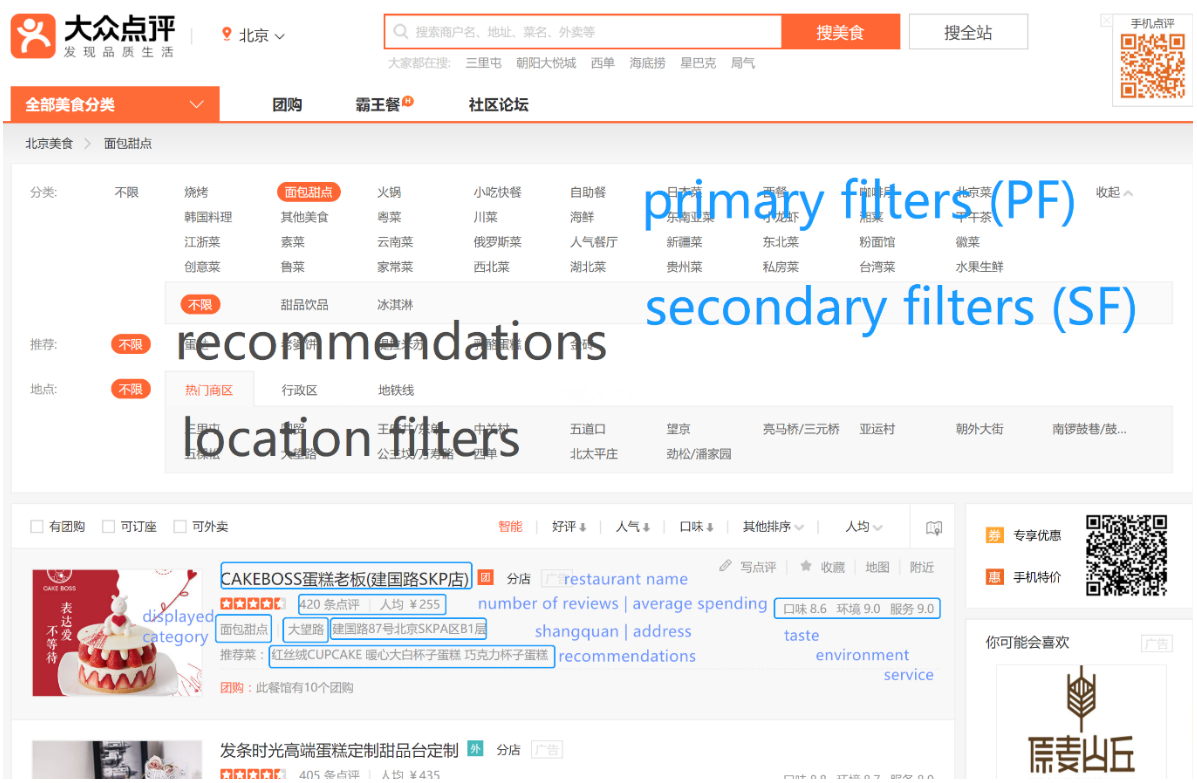


Figure 2 Dianping Website Structure http://www.dianping.com/beijing/ch10

The thesis first extracted the text data of interest into a database relation of eight attributes (Table 2), with the underlined two attributes uniquely determining each row of the data. Next, for geocoding and data analysis in the later steps, additional fields were created and updated in the csv files: city (the same value “Beijing” for each tuple), PF (the value of the functioning PF is generated in each tuple), and SF (created if an SF is available and applied on the web page). Examples of updating the additional fields are shown in Table 3. The PF values for the “others” cuisine category are updated to reveal the actual cuisine type (Table 4).

Table 2 Schema for Data Extraction

Restaurant Name	Address	Average Spending	Displayed Category	Shangqu	Number of Reviews	Recommended Dishes	Taste	Environment	Service

Table 3 Examples of Adding and Updating Attributes (shaded)

Restaurant Name	Address	…	City	PF	SF
			Beijing	hot pot	fish
			Beijing	hot pot	fish
			…	…	…

Restaurant Name	Address	…	City	PF
			Beijing	barbecue
			Beijing	barbecue
			…	…

Table 4 Updating PF for the "others" Type

Original PF	Original SF	Updated PF
others	Shanxi cuisine	Shanxi cuisine
others	mixed	mixed
others	Jiangxi cuisine	Jiangxi cuisine
others	Inner Mongolian cuisine	Inner Mongolian cuisine
others	Tibetan cuisine	Tibetan cuisine
others	other Chinese	other Chinese
others	teahouse	teahouse
others	pub	pub

Based on the restaurants addresses, the csv files were converted into geographic features (WGS 1984) in QGIS using Google Maps as the geocoder. QGIS recorded in a new csv file the addresses which the geocoder failed to locate (the “not found” addresses). Initially, 33.44% of the total 34485 addresses were not found by Google Maps. This is mainly because many restaurants provide too much detail (Table 5) or simply described their locations rather than providing a formal address; for example, X meters south from the intersection of Y Street and Z Road.

In lack of available alternative geocoders, all the “not found” addresses were manually edited. To understand what forms of address is acceptable by the Google geocoder, the “not found” addresses in the barbeque data set were edited and searched on the Google Developers (2018) web page. Successful retrieval of map information on this web page means successful geocoding by Google Maps. General rules of manual editing are listed as follows, and a typical example is given in Table 5.

a) Convert the address column in Excel, delimited by “号” (a symbolic character following a number);
b) Use district name rather than any other levels of administrative division;
c) The best bet is to delete everything after the delimiter and edit the part of address before it (Edition 1).

Table 5 An Example of Manual Editions

	Administrative division	Street	Delimiter	Building	Room
Original address	黄村	金星路 3	号	缤纷城	东区 3 楼 301
Translation by semantic unit	Huangcun Village	3 Jinxing Rd		Binfencheng (informal name of a mall)	301, 3rd floor, East Wing
Edition 1	大兴区	金星路 3			
Translation	Daxing	3 Jinxing Rd			
Edition 2	大兴区			绿地缤纷城	
Translation	Daxing			Lvdi Binfencheng (full name)	

To edit the address value that is a description of a location, Baidu Map was used as its database is better maintained in Mainland China than Google Maps. The steps are listed below.

1) Delete details in each address as above and search it in Baidu Maps;
2) Click on a nearby map entity and use its address for geocoding this location.

When all the “not found” addresses were edited, they were geocoded in QGIS using Google Maps. The new “not found” rate was calculated in Excel (see Appendix A) and the overall “not found” rate dropped remarkably from 33.44% to a minimal 2.24%.

All the point features geocoded in the first and the second time were loaded and merged into one shape file. In QGIS, this was completed by the plug-in Load Them All and the tool Merge Vector Layers. Next, the merged dataset was clipped by the boundary of Beijing to delete the points that were geocoded outside the study area. Finally, duplicate rows were removed based on the primary keys (underlined in Table 2) because a restaurant can appear in the search results of multiple filters. This was completed by ArcMap tool called Delete Identical. Typos were corrected in the manually added fields.

6.1.2 Public Data

Table 6 describes the public data and provides their sources. These spatial and statistical data datasets are free for the public to download.

Table 6 Public Data

Data Description	Sources
Administrative divisions in Beijing (polygons)	GADM www.gadm.org
Population by district (text)	Beijing Statistical Yearbook 2017 http://www.bjstats.gov.cn/English/AD/
Expressways in Beijing, 2012 (polylines)	Beijing government open data (shp) http://www.bjdata.gov.cn/

6.2 Describing the Overall Point Distribution

The first step was to describe the point pattern on the city level and the district level. All the restaurant points were mapped and visually examined on the city level. For the 16 administrative districts, the count and the density of restaurants w ere compared to the population and population density of each district. In ArcMap, restaurant count and density were obtained by spatially joining the restaurants points to each of the district polygons that completely contains a subset of the data points. A new attribute called Join_Count was created, and the values equal to the count of restaurants within each district. Restaurant density was calculated in a new field by dividing Join_Count by the area of each district polygon calculated by Calculate Geometry. Population density was calculated by dividing the number of permanent resident by 1 km^2.

Secondly, the entire point pattern was characterized centrographically. The mean center of the point pattern was compared to the geometric center of the city. The ArcMap tool Mean Center created the mean center point feature using the mean coordinates. The dispersion and directional trend of the entire dataset was characterized by standard deviational ellipses using the Standard Deviational Ellipse tool. A way to understand how the restaurants are clustered around the mean center is to calculate the percentage of the restaurant points that fall within each standard deviational ellipses. This was done by Spatial Join in ArcMap.

Although Moran's I is commonly used to indicate global autocorrelation, it is not designed for describing density of point incidents. Moran's I measures attribute values across locations, and therefore requires the input point dataset to have a field of values to report about. In ArcMap, Collect Events was used to assign weights to the point pattern, and the Spatial Autocorrelation tool to measure Moran's I statistic for the weights at each location. Inverse distance was used as the parameter to account for the fact that nearby features has greater influence on a location than faraway features, and Euclidean distance was considered because the street network allows flexible route options for customers to travel to restaurants.

6.3 Identifying Local Clusters

Getis-Ord Gi or local Moran's I can be used to identify statistically significant clusters. Since these methods consider the cluster of high or low attribute values of a feature dataset, a quadrat count was first performed, and the quadrats containing a significantly large number of restaurants were identified using the statistical measures.

The ArcMap tool Create Fishnet was used to create quadrats over the extent of the study area. It is crucial to determine an appropriate quadrat size because using large quadrats will generalize local variations and using small quadrats will substantially increase the number of low count values. The thesis experimented with the parameters for the fishnet, and an appropriate specification is to divide the rectangular extent of Beijing into 400 columns and 400 rows. This determined the cell size to be 445m horizontally and 448m vertically. The cell size was considered appropriate because it is similar to the size of a street block in Beijing, measured on satellite image by the ArcMap tool Measure. The restaurant points were spatially joined to each quadrat that completely contains them, and the join count was recorded as a new attribute. To visually examine the restaurant density represented by quadrat count, the quadrats were classified into 5 classes using the natural breaks method which maximizes the differences between the classes.

The Hot Spot Analysis in ArcMap was used for the quadrat features containing join count values. In Section 7.3.1, the result shows that many hot-spot quadrats are clustered. Standard deviational ellipse is used to characterize the clusters of significant quadrats. The identified hot spots (confidence level ≥ 90%) are exported as a new shapefile to construct standard deviational ellipses. The steps to select the members of a hot-spot quadrat cluster are provided in Section 7.3.2 as it is easier to understand the process with the images.

7. Results

7.1 Data Collection and Preprocessing Results

33714 (97.76%) of 34485 scraped data were successfully geocoded. Of the 33714 geocoded data points, 70 (0.2%) point events were located outside the study area and thus removed from the data set. Of the remaining data points, 1280 duplicate data entries were deleted based on the primary key (Table 2). The main reason for duplicate restaurant is that a restaurant is allowed to appear in the search results of different filters on Dianping.

Finally, 32802 unique restaurants were mapped within the study area. The data set is considered desirable for its large size and the minimal levels of errors in the geocoding process. Figure 3 is the map of all the restaurant points with a zoom-in view of the central districts. Many parts of the city are almost blank while a few streets in the central area are noticeably delineated by the points.

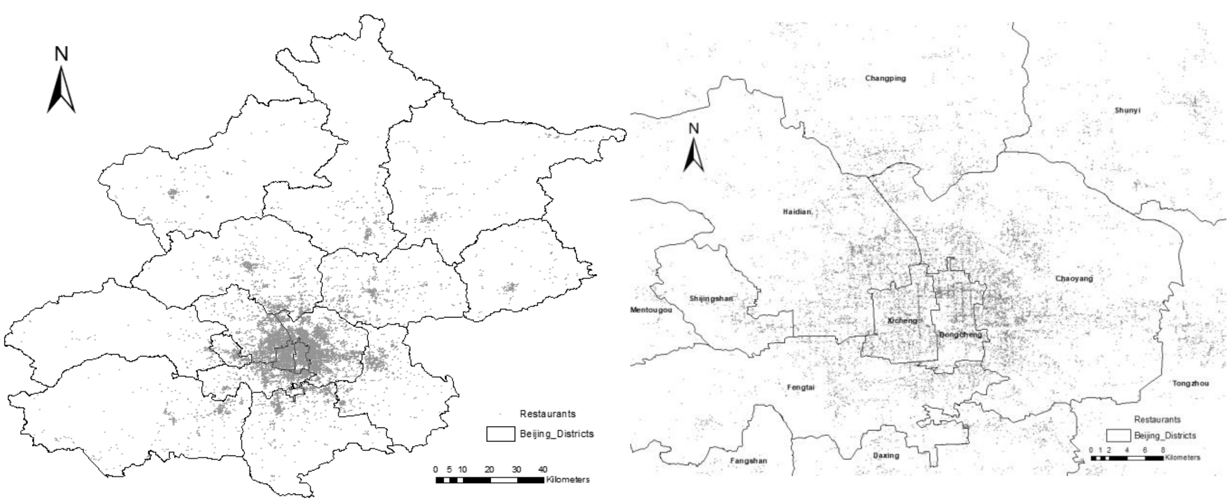


Figure 3 Mapping the Restaurant Points

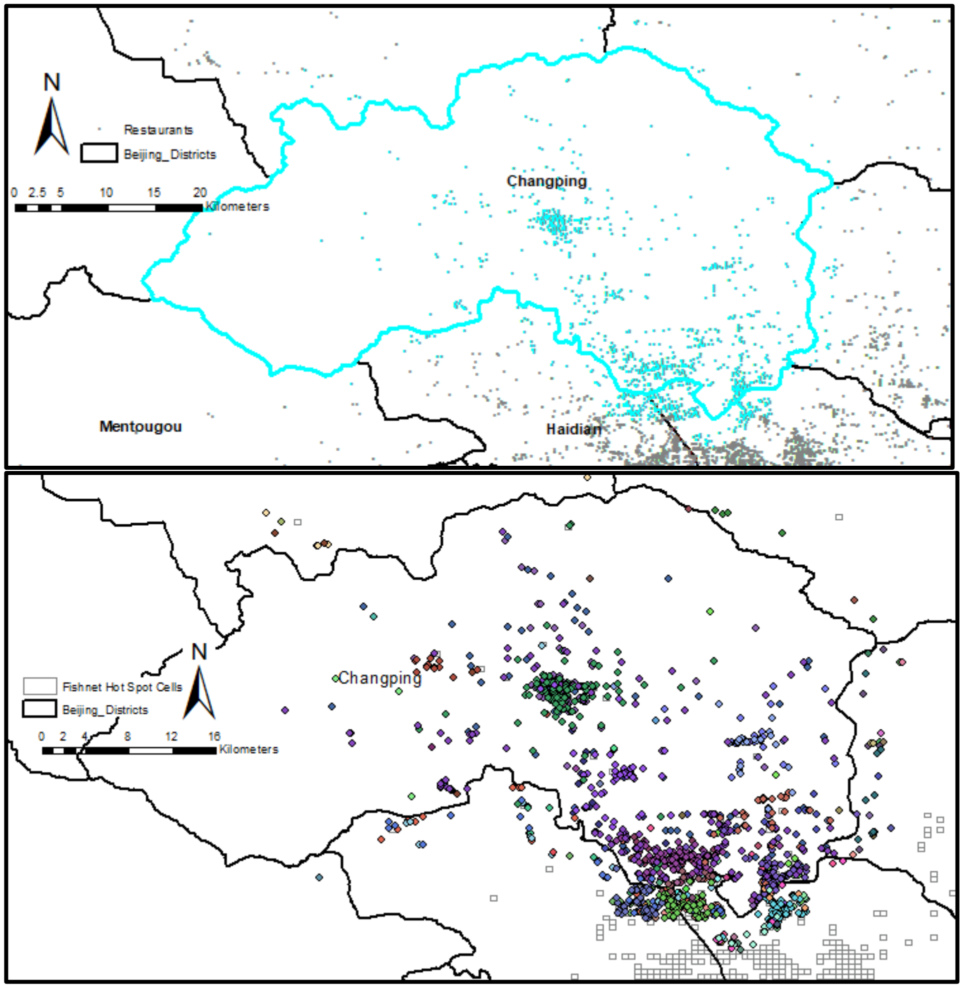


Figure 4 Classifying Points by Shangqu Values

The restaurant dataset contains a shangqu attribute provided by the restaurants. Referring to Figure 4, restaurants points in Changping District and within 3 kilometres from the district boundary were selected using the Select By Location tool. Points were classified by the shangqu attribute, presenting restaurants with the same shangqu values in the same colors. This attribute is questionable as one location can be given different shangqu values (such as the name of the jurisdiction, the name of a mall, and the name of a historic landmark). This might suggest the boundary of a shangqu varies across personal perceptions.

7.2 Overall Point Pattern

The first way to examine the point pattern is to look at restaurant count and density by administrative district. The pattern is compared to the permanent population count and density in 2016, which is the most recent data available.

The density of restaurant points follows the pattern of population density (Figure 5). The core of the city, Dongcheng and Xicheng, is the most densely populated area and most densely occupied by restaurants. However, due to the small size of this area, the districts around it are the highest in population and restaurant count. Wile Chaoyang and Haidian have the largest numbers of population, Chaoyang attracts most restaurants

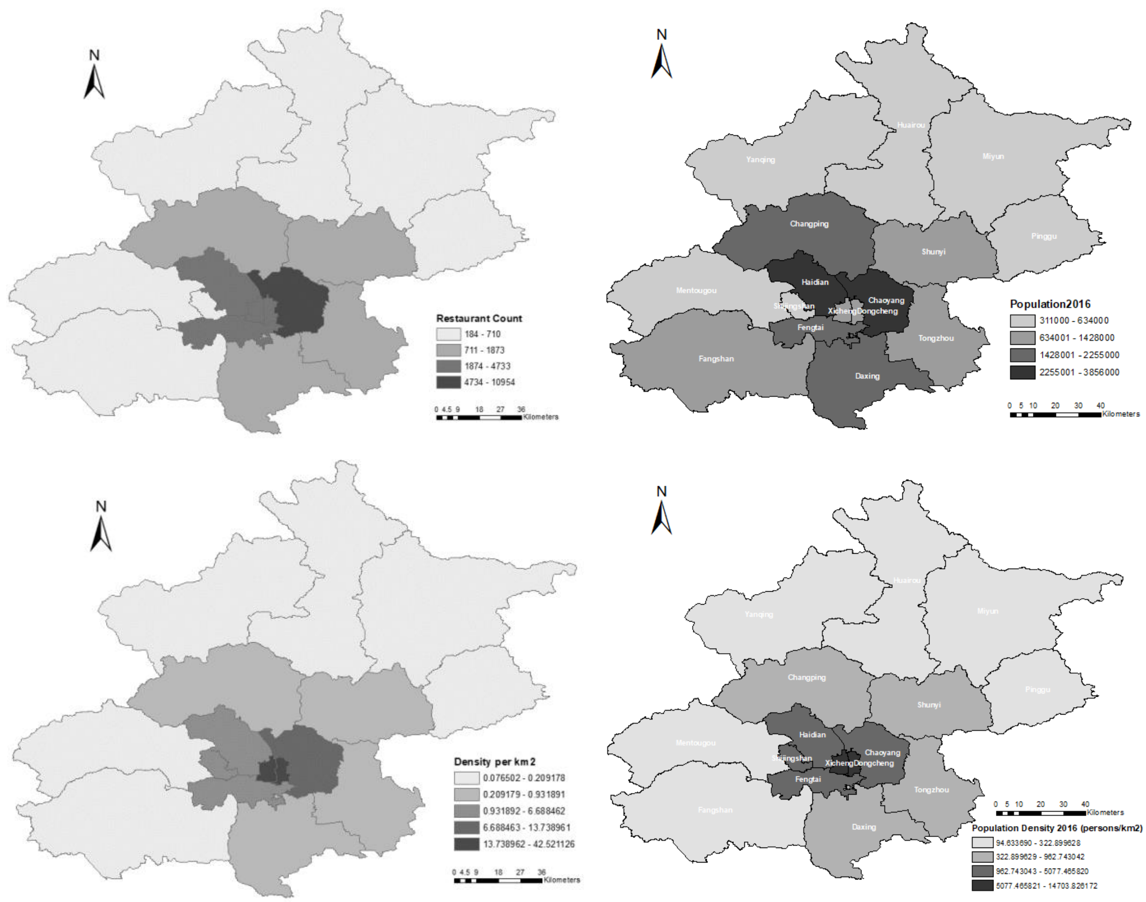


Figure 5 Comparing Restaurants and Population

among all the districts. High restaurant density appears to be spreading northwards and eastwards from the core districts (Dongcheng and Xicheng). Fangshan and the ecological conservation zones are noticeably low in restaurant count and density.

Referring to Figure 3 and 4, it appears that the center of the entire point pattern is the core area, Dongcheng and Xicheng, and that there is a directional trend toward northeast. Centrographical description of the point pattern supports this finding.

Figure 6 shows the centrographical description of the entire point pattern. This approach is not affected by the artificially defined boundaries of the districts. The mean center of all the restaurants (mean coordinate values) is near the Temple of Earth, Dongcheng above 2nd Ring Road North. It is about 12 km south from the geometric center of Beijing in Xiaotangshan County, Changping.

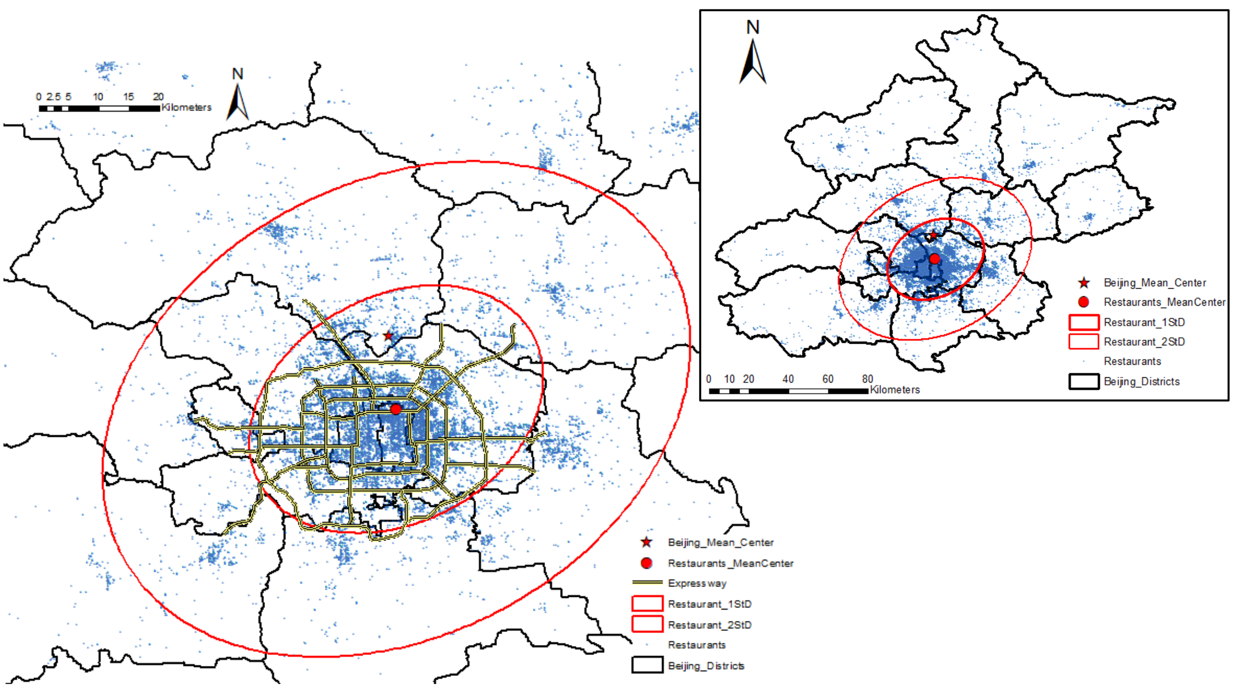


Figure 6 Centrographical Description of the Point Pattern

Standard deviational ellipses characterize the dispersion and directional trends of the restaurants. An N (N=1,2,3) standard deviational (NStD) ellipse is centered at the geometric mean, with the length of its semi-major axis equal to N times the standard distance (average of coordinate differences from the mean in the same direction). The standard distance from the mean center is about 22km (the semi-major axis of the 1StD ellipse) and 17km (the semi-minor axis of the 1StD ellipse). This suggests that more restaurants are located in one direction than another. The ellipses show a 61.78° clockwise rotation from the north, which agrees with the previous finding that there is a directional trend towards northeast from the core area.

Figure 6 visually presents that restaurants are clustered within the 5th Ring Road. The 1StD ellipse contains 25898 (78.95%) of the restaurant points while the 2 StD contains 31014 (94.55%) of all the points. The ArcMap tool Spatial Autocorrelation reports that the cluster is less than 1% likely to be created by random chance (p-value = 0, z-value=10.35), with a Moran's Index (0.014) approximating zero due to the large sample size.

7.3 Local Clusters

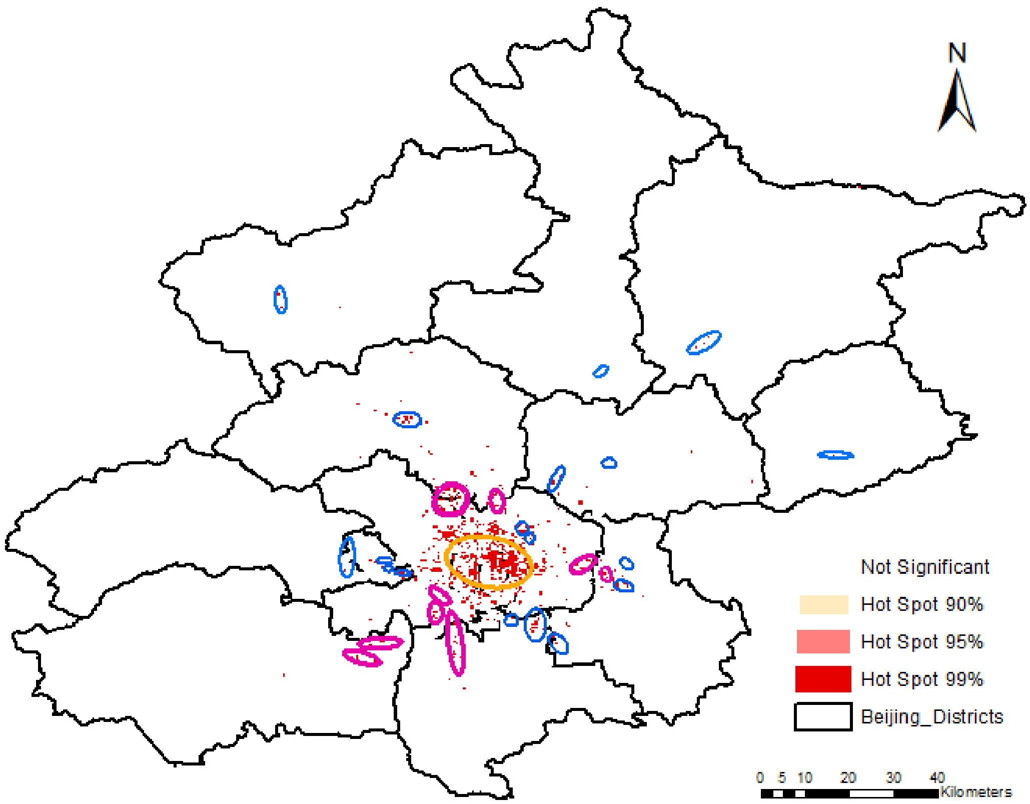


Figure 7 The Major Clusters

Figure 7 shows the result in fulfilment of Research Objective B, the delineation of 28 major restaurant clusters (MRCs). An MRC is a cluster of the hot-spot quadrats identified in Section 7.3.1. Each ellipse characterizes a group of neighboring quadrats of statistical significance. The thesis defines that the boundaries of the ellipses must not intersect. Section 7.3.1 presents the restaurant quadrats of statistical significance while Section 7.3.2 explains the different situations for describing the cluster of the significant quadrats.

7.3.1 Restaurant Quadrats of Statistical Significance

The ArcMap Tool Optimized Hot Spot Analysis identifies statistically significant clusters of high values (hot spots) and low values (cold spots). The tool evaluates the point events to determine the optimal quadrat size for analysis, aggregates the point features by each quadrat, and outputs the high spots and cold spots using the Getis-Ord G*i** statistic.

Figure 8 shows the output of the Optimized Hot Spot Analysis. The quadrat size is about 1.3km by 1.3km. The tool is not optimal for the thesis as the quadrat size cannot be determined manually. ArcGIS (2016 c) recommends using the Hot Spot Analysis tool instead to have more control over the process.

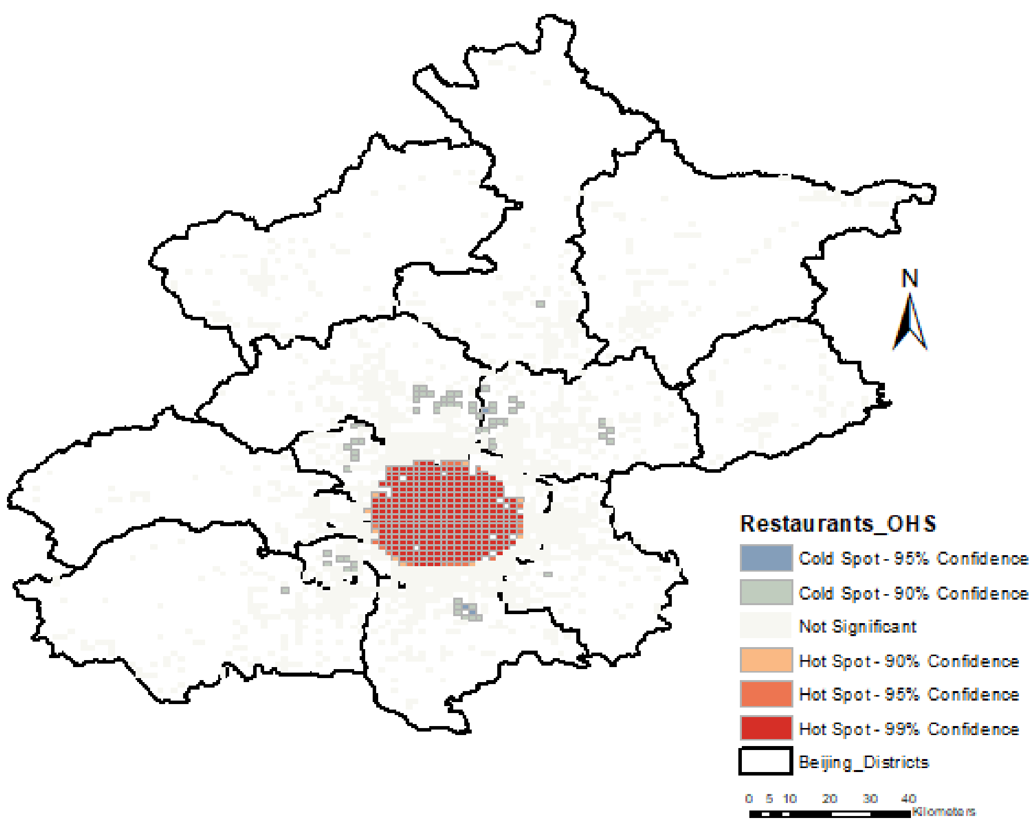


Figure 8 Optimized Hot Spot Analysis

To use the G*i*-based Hot Spot Analysis, each location *i* must have an attribute value to calculate the G*i* values. The thesis created a fishnet over the extent of the study area, clipped the fishnet to within the boundary of Beijing, and performed the hot spot analysis using the number of points within each fishnet cell. An appropriate specification for the fishnet is 400 columns and 400 rows, which determines the cell size to be 445m horizontally and 448m vertically. The cell size is considered appropriate because it is similar to the size of a street block in Beijing. Figure 9 shows the number of restaurants in each fishnet cell (or quadrat).

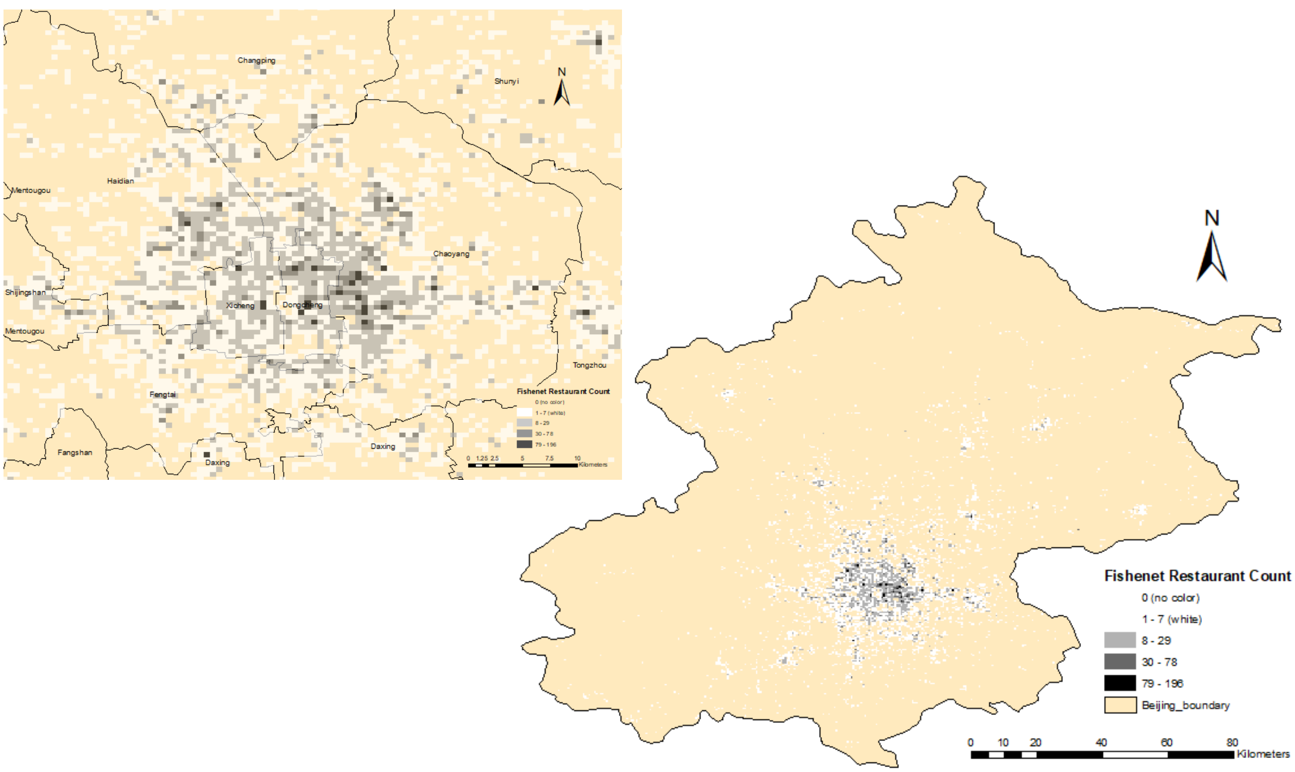


Figure 9 Fishnet Quadrat Count

In Figure 9, the quadrats are classified into 5 classes. 78199 (93.69%) out of the 83463 quadrats have no restaurants within them. The other quadrats are classified into 4 classes using natural breaks that maximize the differences between classes. Table 7 shows the distribution of restaurant count values.

Table 7 Distribution of Restaurant Count Values

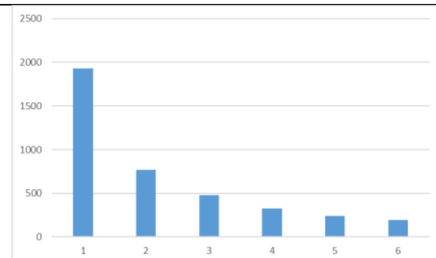
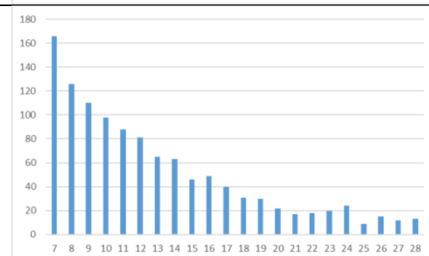
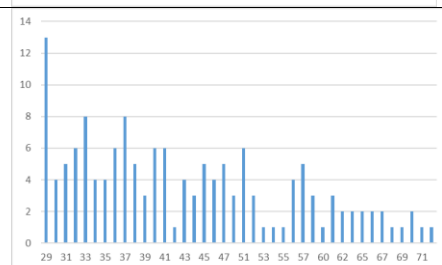
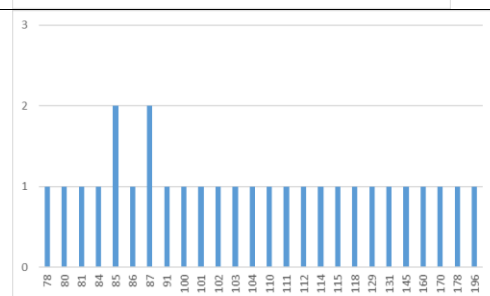
Restaurant count (x)	Total quadrats in this class	Number of fishnet cells (y) for each x
0	78199	
1-6	3946	
7-28	1143	
29-77	147	
78-196	28	

Figure 10 shows the result of the hot spot analysis based on the fishnet quadrat count. Different colors and symbol sizes are applied to visualize the confidence levels of the hot spots. Each of the colored quadrats indicates a statistically significant (based on Getis-Ord G*i**) local cluster of restaurant. It is noted that the hot-spot quadrats show clustering patterns. The clusters of hot-spot quadrats are characterized by standard deviational ellipses (Section 7.3.2).

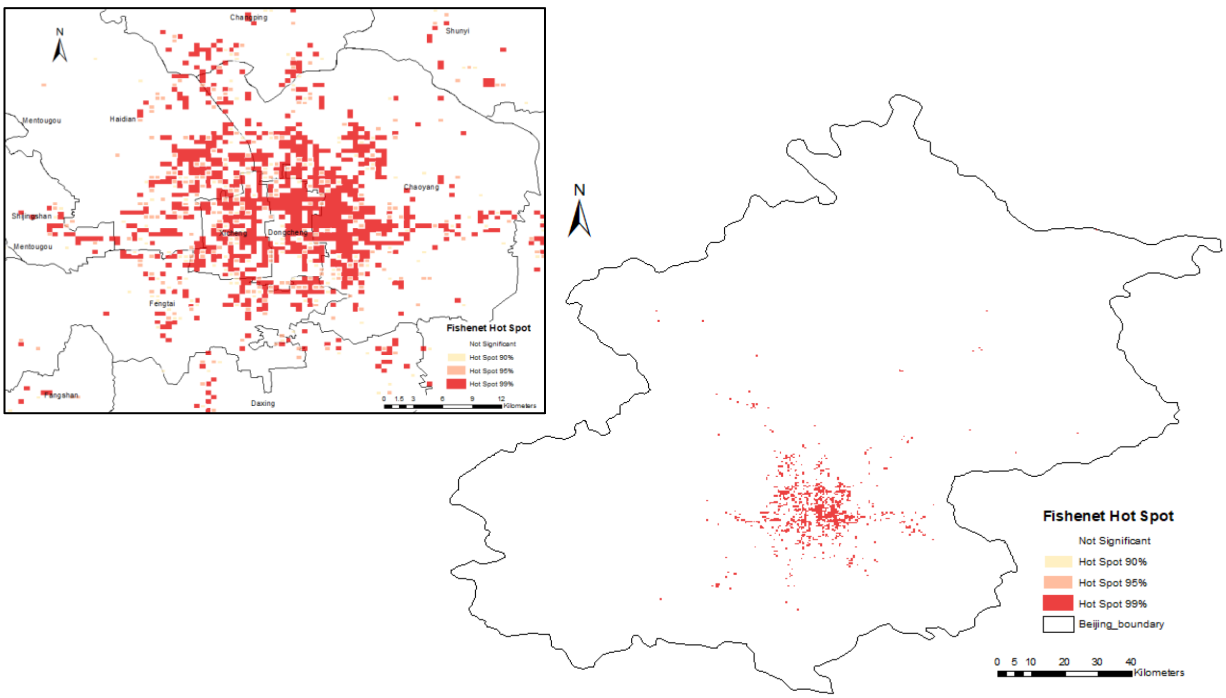


Figure 10 Result of Hot Spot Analysis

Figure 10 shows the result of the alternative measuring tool for local clusters, Cluster and Outlier Analysis. Using the same input as the hot spot analysis, Cluster and Outlier Analysis outputs statistically significant (based on Anselin Local Moran's I) quadrats of the following 4 types (LMI types).

1) LL: quadrats containing a significantly small number of restaurants surrounded by similar quadrats (not existing in the study area)
2) LH: quadrats containing a significantly small number of restaurants surrounded by dissimilar quadrats
3) HL: quadrats containing a significantly large number of restaurants surrounded by dissimilar quadrats
4) HH: quadrats containing a significantly large number of restaurants surrounded by similar quadrats

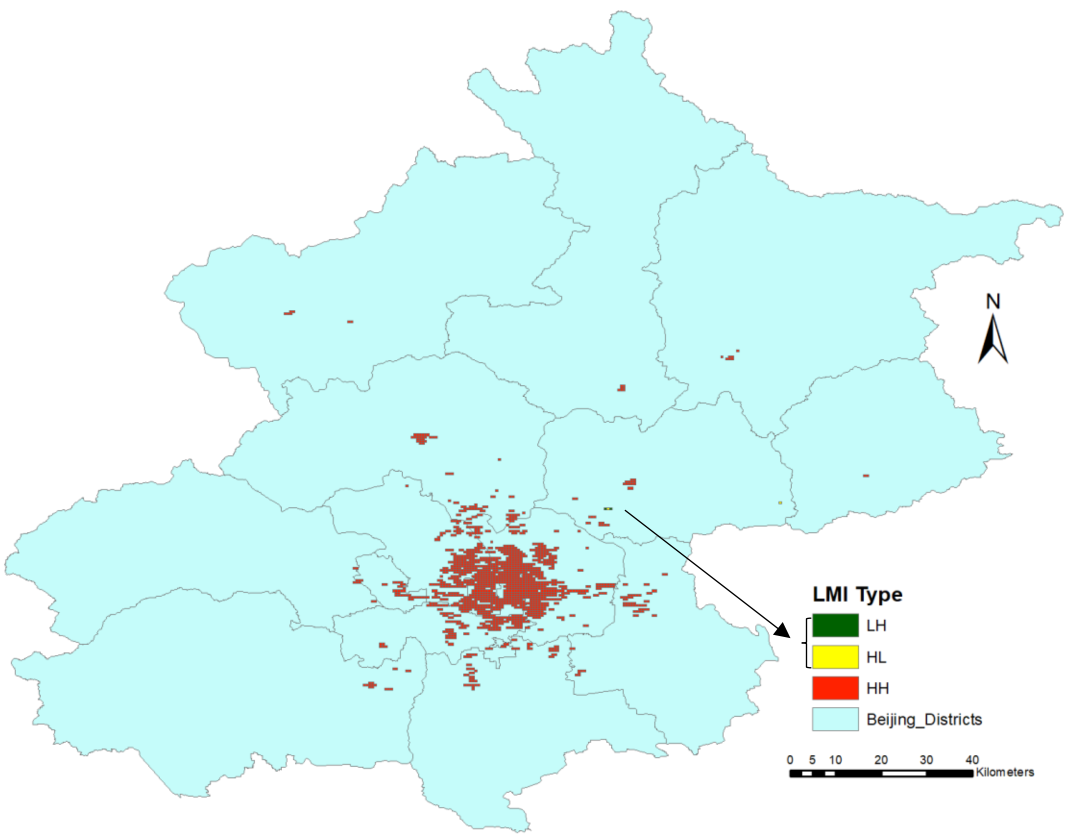


Figure 11 Result of Local Moran's I

Referring to Figure 11, HL and HH quadrats are the local clusters the thesis is interested in. It is noted that there are only 1 HL quadrat and 2 LH quadrats in the results. Similar to the hot spot analysis result, the HH quadrats also show clustering patterns.

An HH quadrat seems similar to a hot spot in the hot spot analysis. However, Getis-Ord G*i* determines the study area's hot spots (comparing the value to the entire study area) while an HH quadrat is significantly similar to its high-value neighborhood. Therefore, the thesis considers the result of the hot spot analysis which focuses on significantly high values over the study area and provides the confidence levels for the hot spots.

7.3.2 Clusters of Significant Quadrats

In Section 7.3.1, both the hot spots and the HH quadrats show clustering patterns. The thesis defines such a cluster as a major restaurant cluster (MRC). As mention above, each MRC is characterized (or delineated) by a standard deviational ellipse.

A group of neighboring significant quadrats are selected, and an ellipse will be constructed based on their coordinates. Instead of using the HH quadrats, the thesis considers selecting the hot-spot quadrats as their confidence levels are provided and sufficiently high (≥90%). Sections 7.3.2.1 to 7.3.2.3 explain the different situations for selecting the members of an MRC.

When using the Directional Distribution tool in ArcMap, only 1 standard deviation or 2 standard deviation was applied. This is to ensure that the ellipses do not intersect with one another or cover an overly large area. The number of restaurants within each hot-spot quadrat (the Join_Count attribute value) was used as the spatial weight to construct the ellipse.

7.3.2.1 Isolated Major Restaurant Clusters

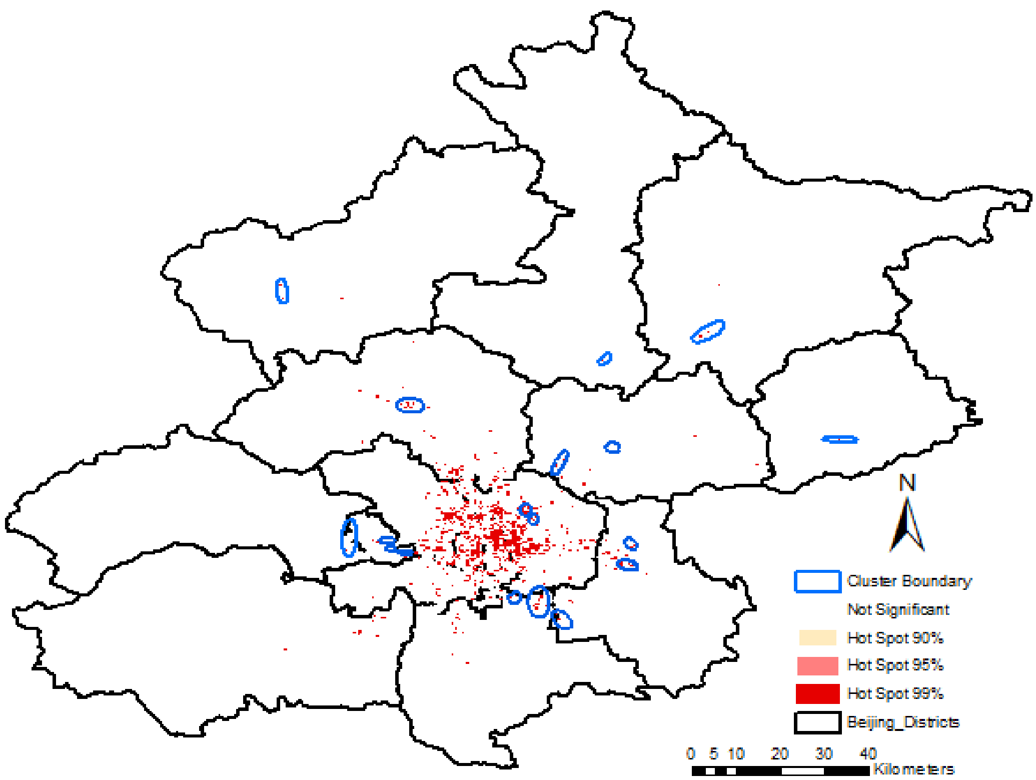


Figure 12 Isolated Major Restaurant Clusters

Referring to Figure 12, 18 major restaurant clusters (MRCs) were easily identified because the members of each cluster are close to each other and sufficiently far away from other hot-spot quadrats. Isolated MRCs are more common in the administrative districts that are low in population density. Ten of them used to be the counties (a lower level of administrative division than city in China) before merging into

Beijing’s city boundary, while the others are located near the counties or the core of the city (Dongcheng and Xicheng).

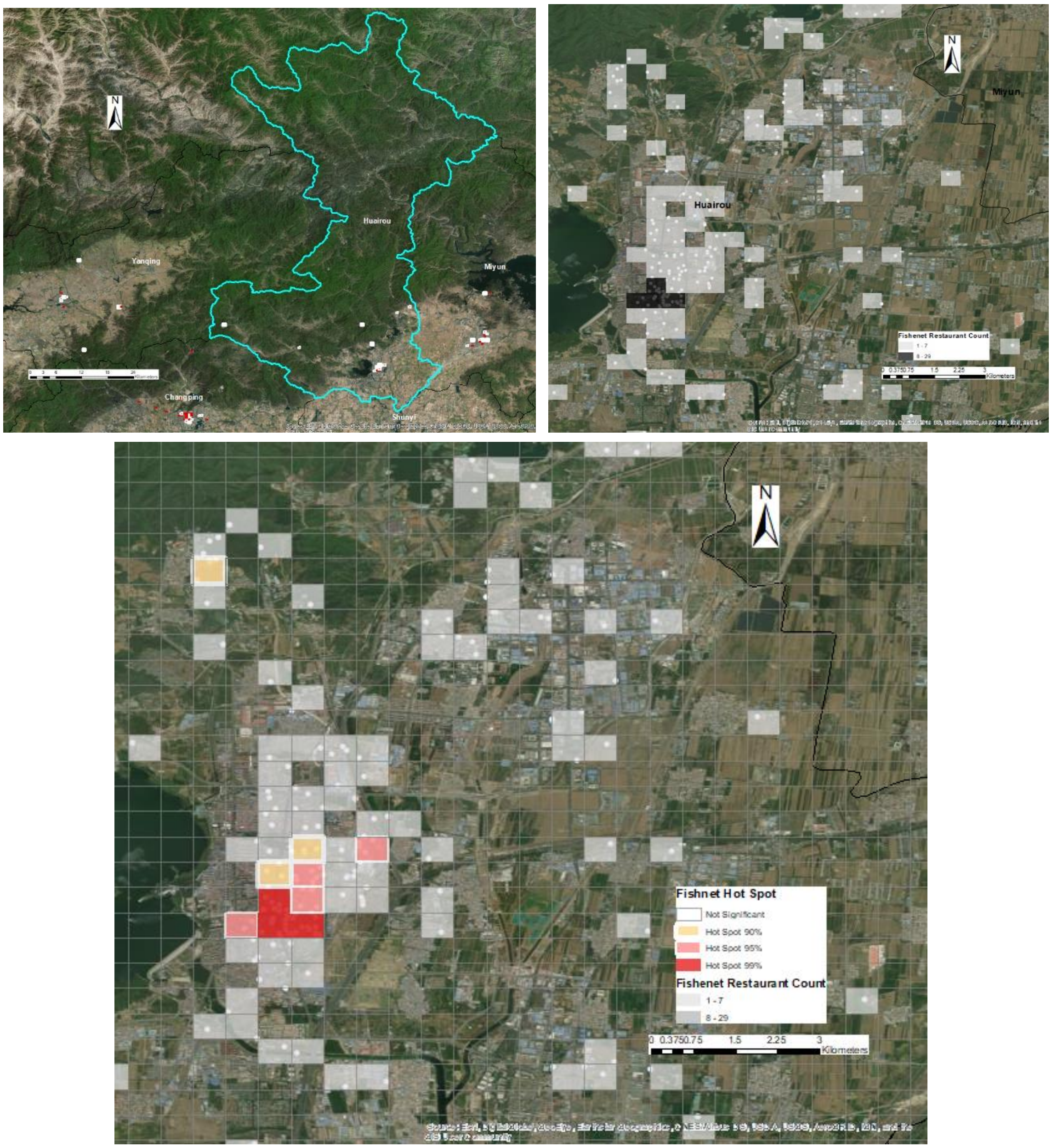


Figure 13 The Hot Spots in Huairou

Figures 13 and 14 exemplify the process of selecting the MRC members and constructing the standard deviational ellipse in Huairou. Referring to Figure 12, only one location in Huairou is occupied by more than one hot-spot quadrats. The quadrat count layer and a satellite base map are used to select the appropriate MRC members. Compared to the classified quadrat count values, the G*i* statistic remarkably narrowed down the quadrats to be considered. The 90% hot spot in the northwestern corner is disregarded because the quadrat is

1) Far from other MRC candidates (a quadrat is 445m wide 448m long);
2) Separated from other MRC candidates by quadrats that contain 0 restaurant;
3) Separated from other MRC candidates by a different type of land cover.

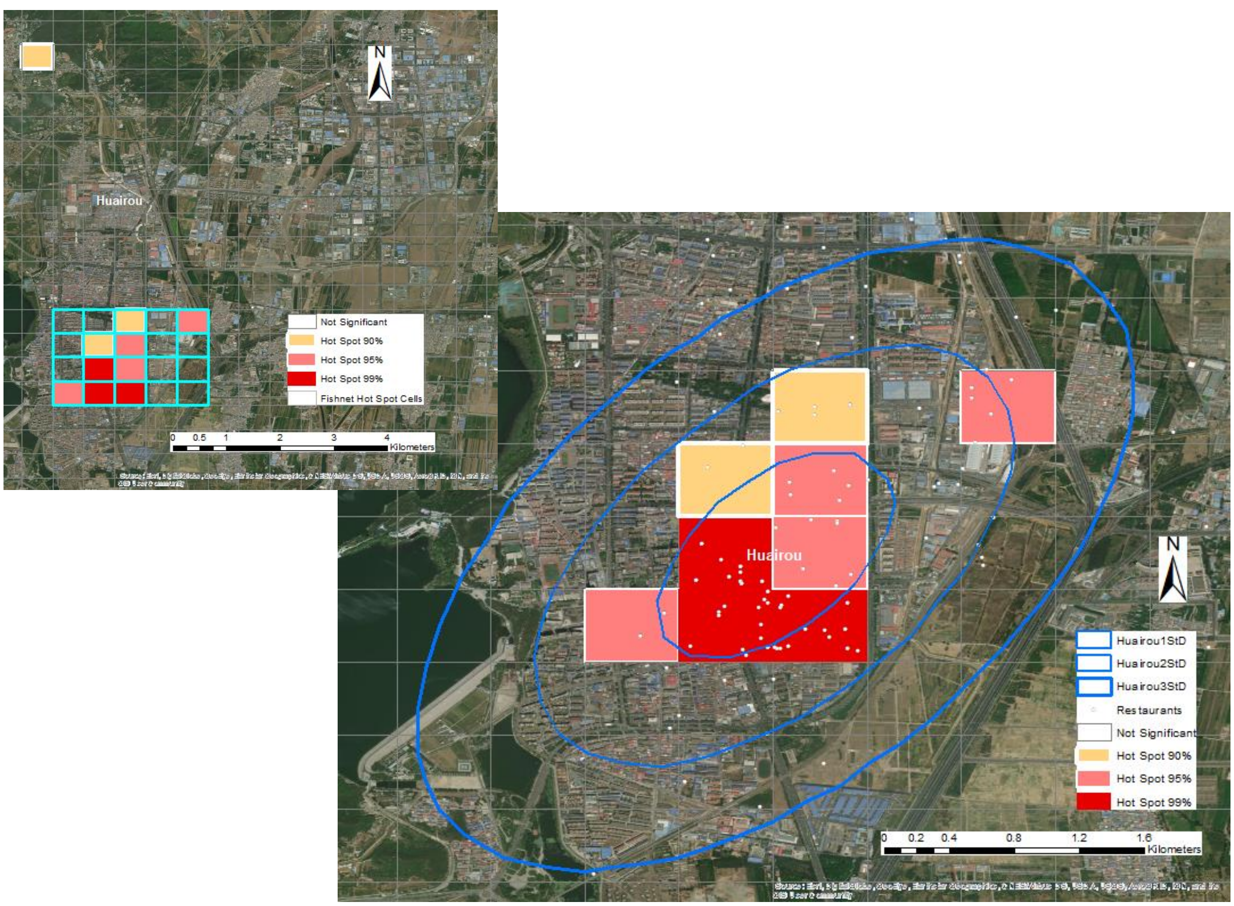


Figure 14 Characterizing the Major Restaurant Cluster in Huairou

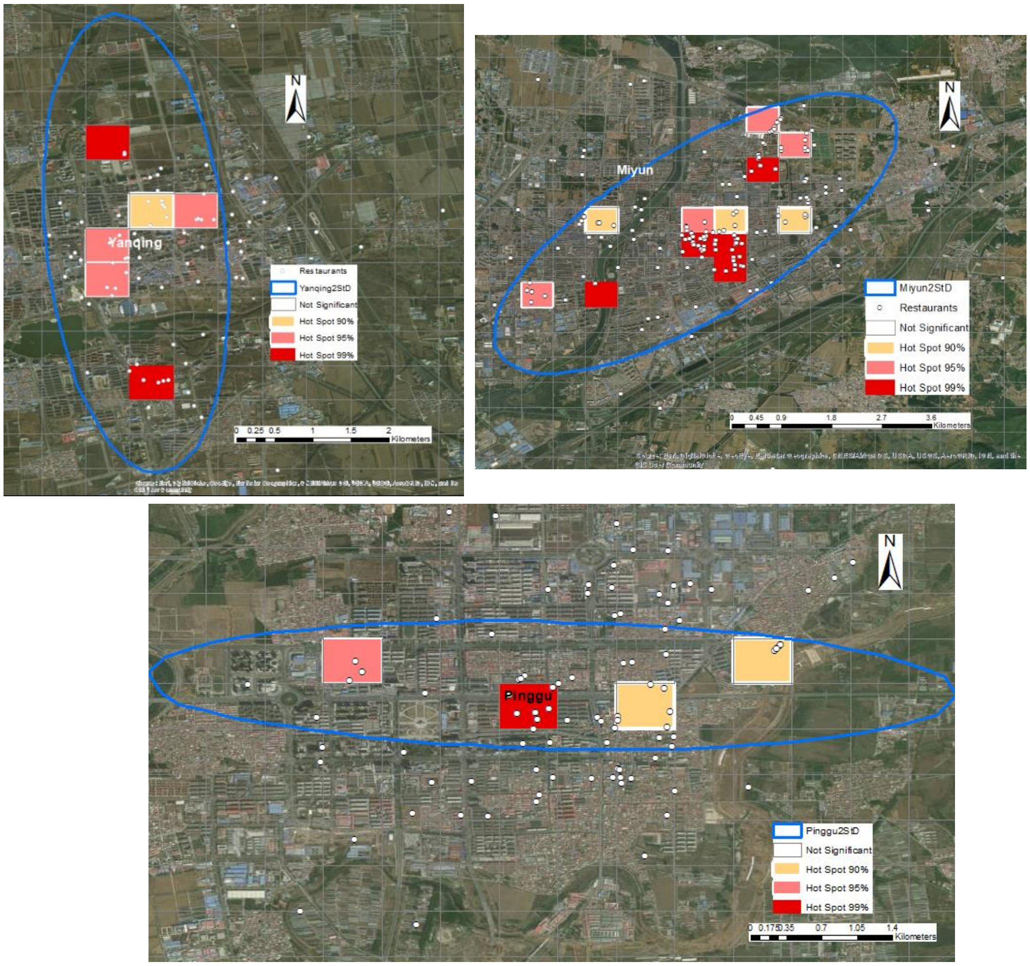


Figure 15 More Examples of Characterizing Isolated MRCs

These are the criteria for excluding MRC candidate quadrats. Referring to Figure 14, the appropriate MRC candidates are selected using Select Feature in ArcMap. In fact, the quadrats in another layer that only contains hot spot quadrats are selected. The selected quadrats of this layer are input to the Directional Distribution tool, and the Join_Count attribute (the number of restaurants within the quadrat) is used to generate spatial weight. In this example, the 2StD ellipse is the optimal delineation because

1) The 2StD ellipse usually contains about 90% points in a clustered point pattern. The 1StD ellipse contains about 70%, while the 3StD ellipse is overly large without significantly improving the coverage percentage.

2) In this example, the 1StD ellipse appears to exclude parts of the 99% hot spots and the restaurant points in them. Referring to the base map, the 3StD ellipse contains irrelevant types of land use: a part of a water body and parts of agricultural lands.

Figure 15 gives more examples of selecting MRC members and constructing the standard deviational ellipses. The criteria above were applied, and 1StD ellipses were used for some of the isolated MCRs to avoid boundary intersection. One interesting finding is that the MRC in Mentougou District is crossing the administrative boundary with Shijingshan District (Figure 16).

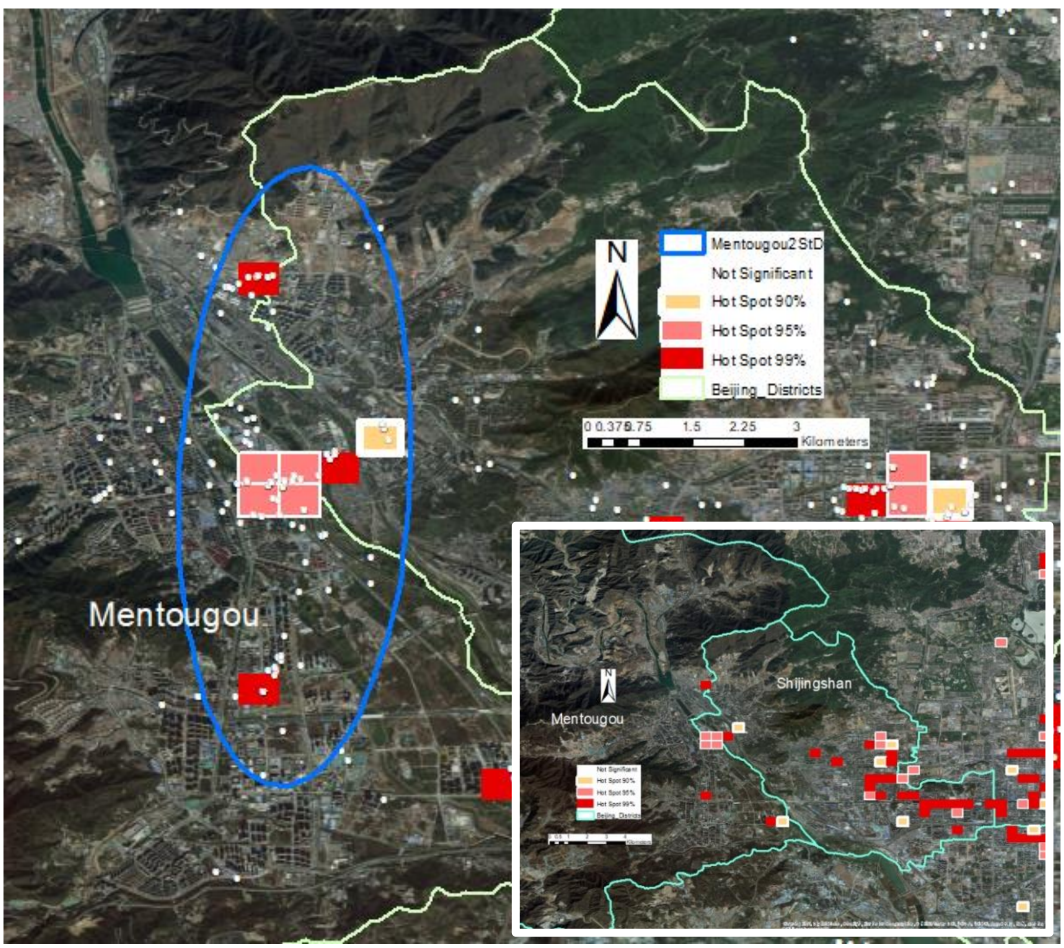


Figure 16 The MRC in Mentougou

7.3.2.2 Combined Major Restaurant Clusters

In areas that are more densely occupied by restaurants, the approach mentioned in 7.3.2.1 often produces intersecting ellipses even when using 1 standard distance. In this situation, the MRC members in the intersecting ellipses are combined to construct a larger ellipse. Figure 17 shows the MRCs that were created this way.

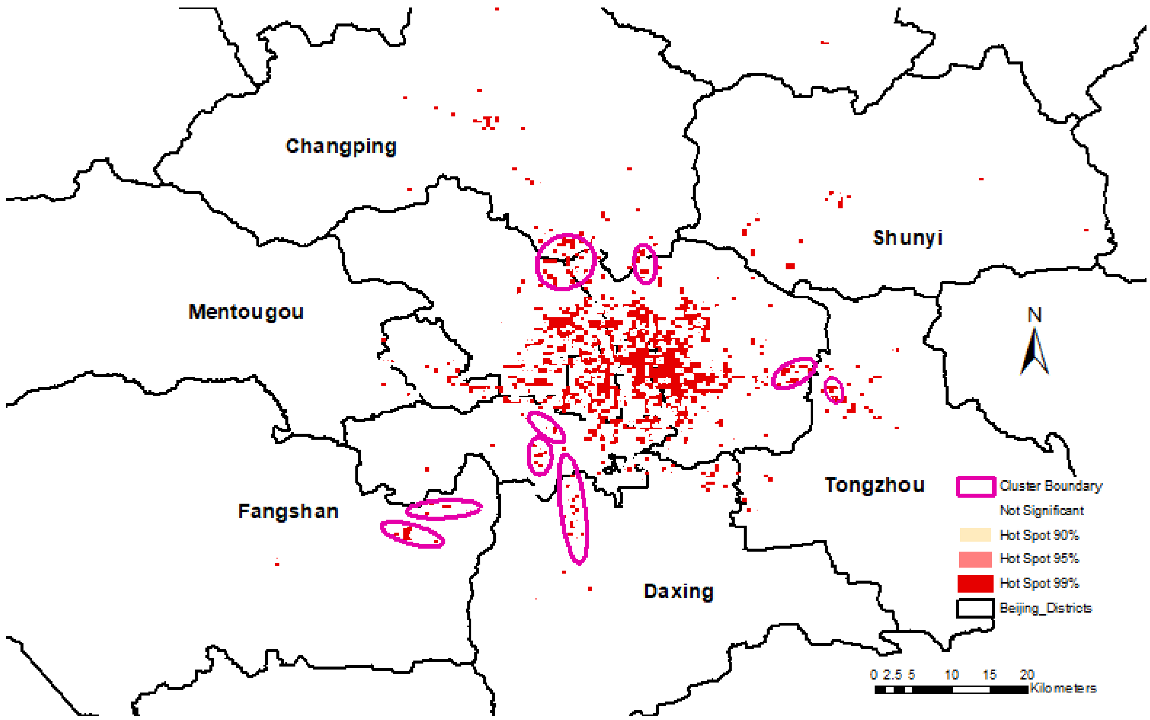


Figure 17 Combined MRCs

This section provides the example of identifying the MRC in Changping District. Referring to Figure 18, one isolated MRC at Changping County was identified, but it is more complex to select the MRC members in the district's southeastern corner as they are close to the hot spots in the bordering districts.

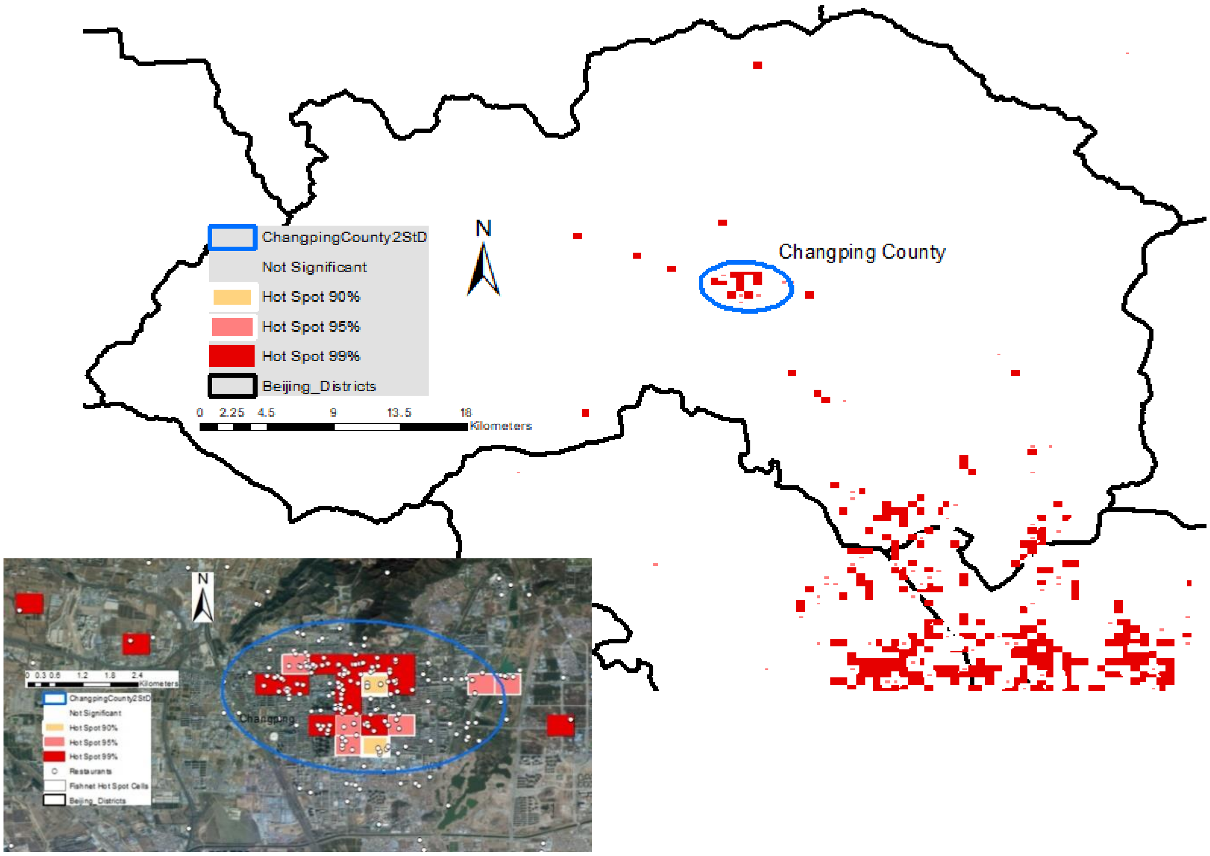


Figure 18 Hot Spots in Changping District

Figure 19 shows an attempt to select MRC members within Changping District for the MRC at the Huilongguan Township of Changping District. The Select By Lasso option was used when selecting the features in ArcMap. As a result, the 2 StD ellipse was intersected with the ellipses of nearby MRCs while the 1 StD ellipse could not account for the hot spots at the administrative border. Therefore, administrative boundaries are not considered when identifying the members of an MRC.

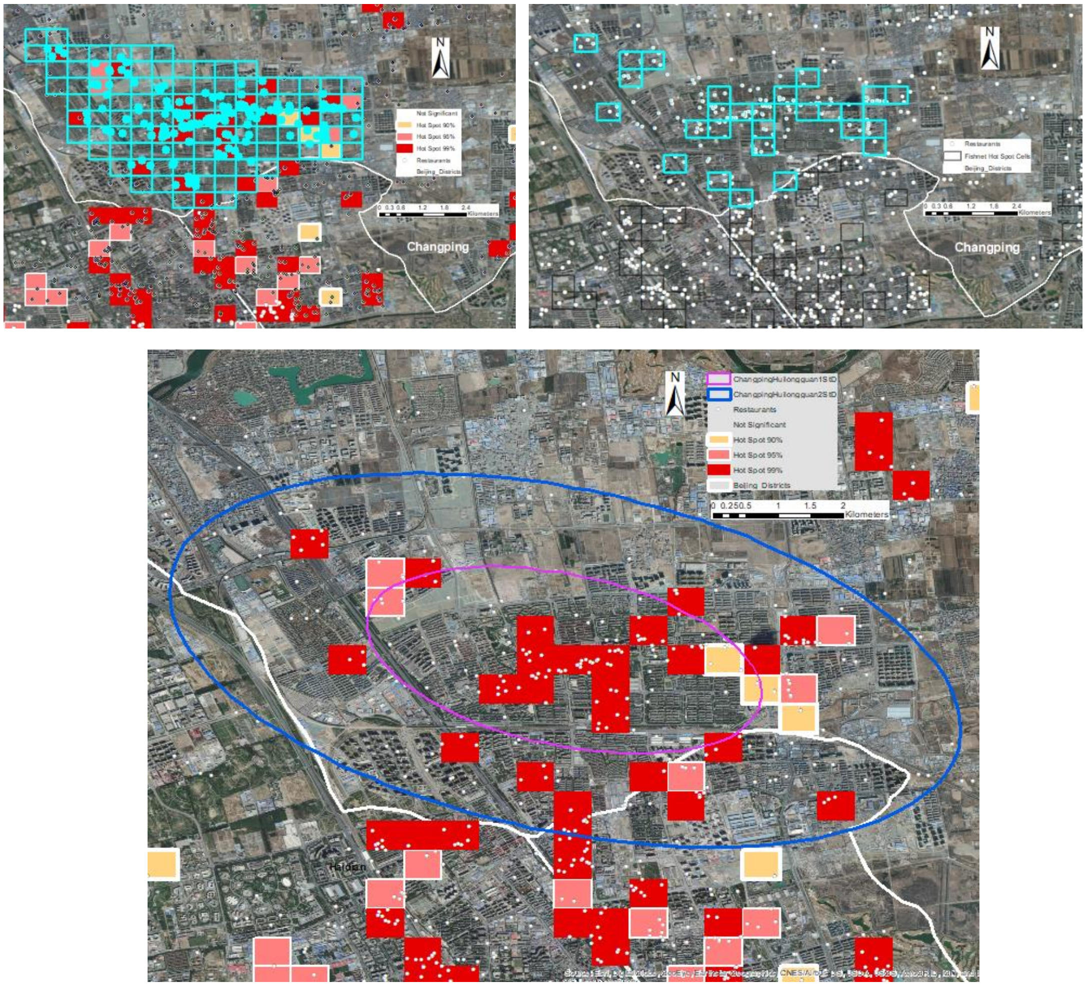


Figure 19 Hot Spots at Huilongguan Township, Changping

Figure 20 shows the MRC members selected and the standard deviational ellipses of them. The MRC members are ensuring to be sufficiently separate from other hot-spot quadrats. The 1 StD ellipse is used to characterize this MRC, as the 2 StD ellipse covers an area of many non-selected quadrat and areas containing no hot spots. This MRC is named after the three neighboring districts as as Changping-Chaoyang-Haidian, while some other combined MRC are named after the majority shangqu values in the restaurant dataset.

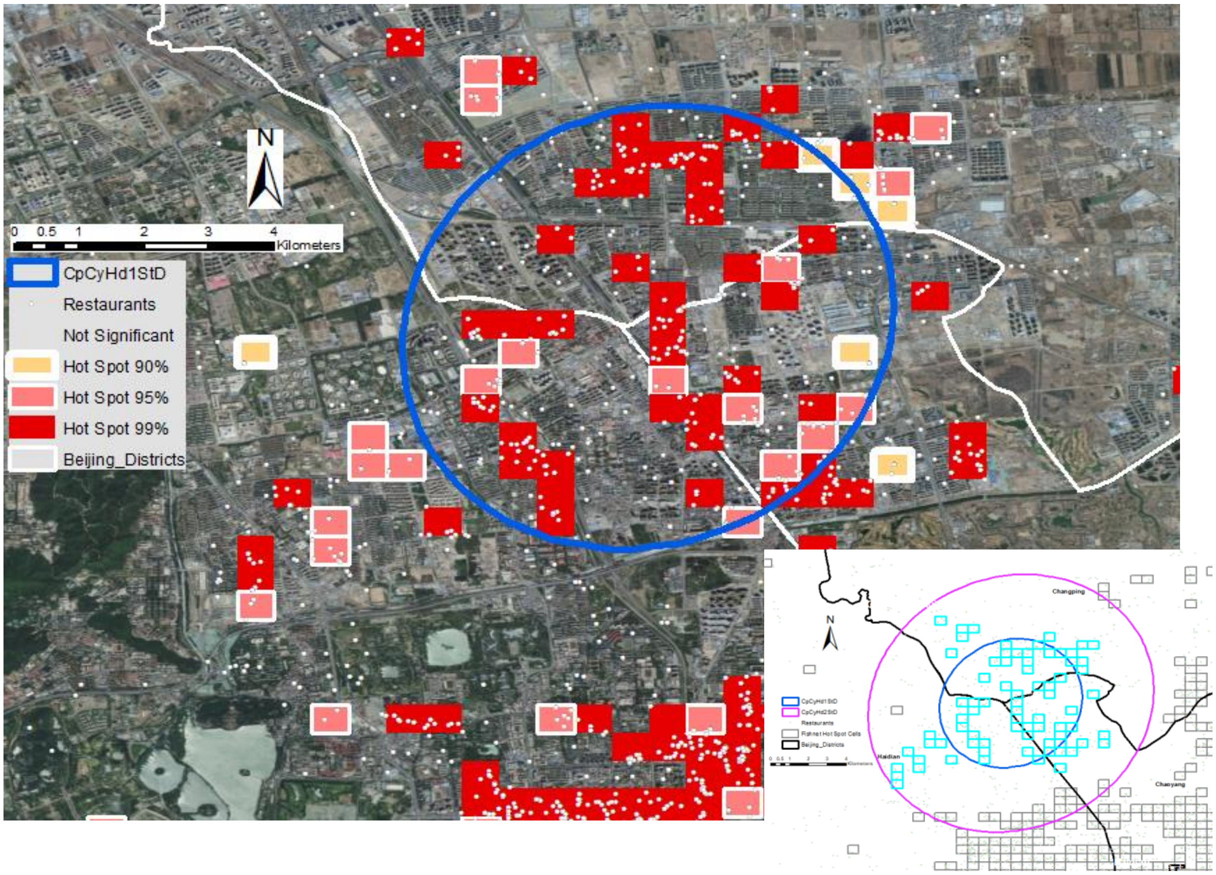


Figure 20 The Changping-Chaoyang-Haidian MRC

7.3.2.3 The Central Cluster

Finally, there is just one big problem left, the large central cluster. The MRC candidates in this area looks inseparable. After delineating all the MRCs around this area (Figure 12 and Figure 17), the hot-spot quadrats in this area appear in the shape of a diamond that is outlined by the blank areas without hot-spot quadrats. Referring to Figure 21, the members of the central MRC are selected in highlight color. Hot spots at 95% and 90% confidence levels are displayed with wide white outlines to show possible subdivisions within the diamond area. The 1 StD ellipse best characterizes the cluster without covering unnecessary blank areas.

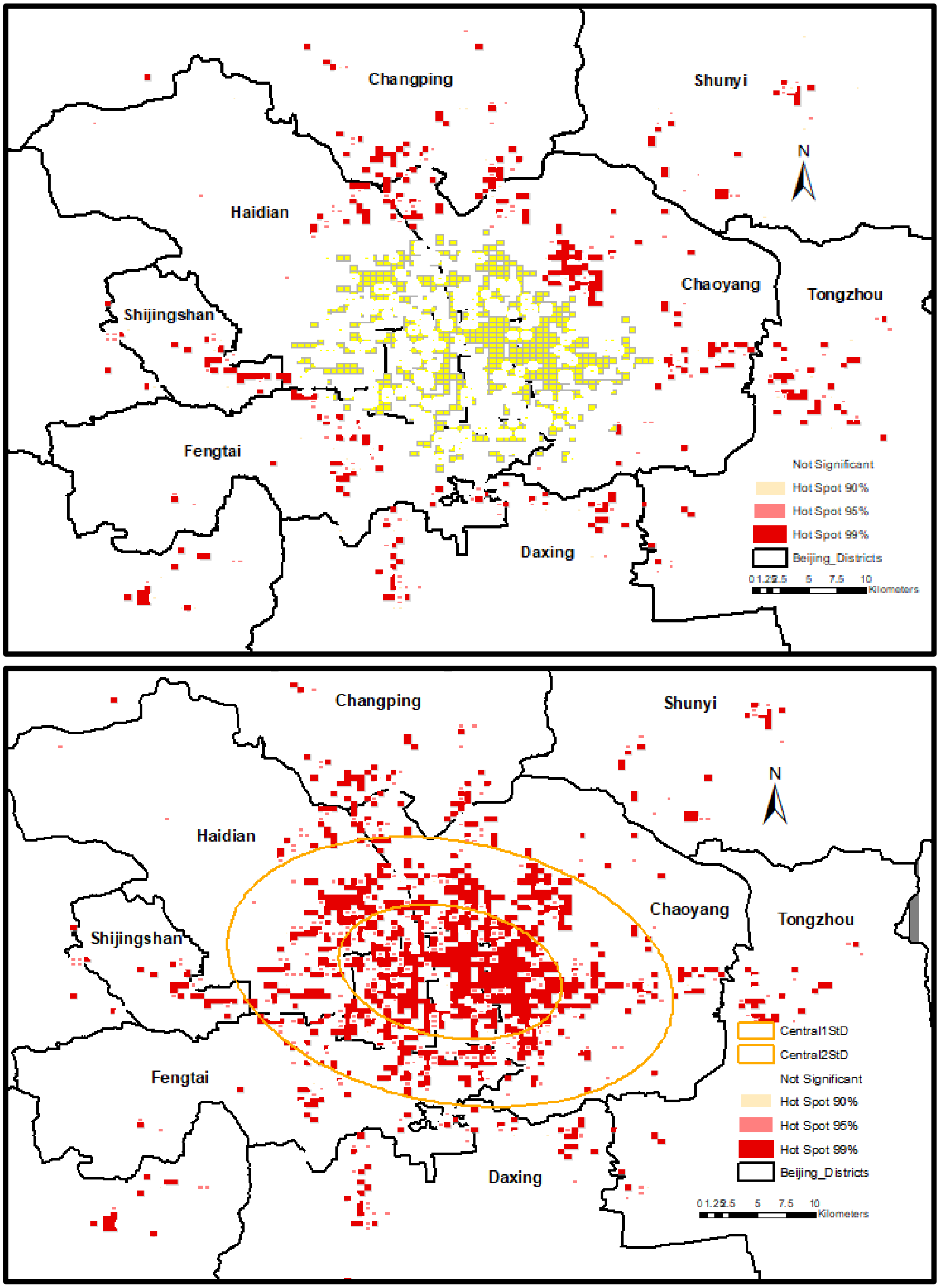


Figure 21 The Central MRC

8. Conclusions

Cluster is a significant spatial pattern, and it can help restaurants make better geographical decisions. The thesis achieves the objectives to describe the global point pattern of the restaurants in Beijing and to identify significant local clusters.

The global point pattern is described by visual examination, comparison to population in terms of count and density, centrography, and Moran's I statistic. The visual presentation of the restaurant points (Figure 3) shows that most parts of the city are almost blank while a visible cluster is centered at the Core Functional Area of the Capital and spread into the Urban Function Extension Area. Restaurant density follows the population density by administrative district, and the northeastern direction from the cluster center attracts more restaurants per squared kilometer than any other directions. The centrally clustering and directionally uneven spatial pattern is confirmed by centrographical statistics. Global Moran's I indicates that the clustering point pattern is not created by random chance with 99% confidence level.

To identify significant local clusters, the thesis utilizes statistical measures based on a fishnet (quadrat size 445m by 448m). Each quadrat is associated with the number of restaurants within it. Based on this value, Getis-Ord G*i* identifies significant clustering high-value quadrats for the entire study area while local Moran's I recognizes high-value quadrats that are surrounded by similar quadrats (Figure 11).

Both the hot-spot quadrats and HH quadrats appear to be clustering. To describe this clustering pattern, the thesis defines a cluster of the hot-spot quadrats as a major restaurant cluster (MRC). Each MRC is characterized by a standard deviational ellipse (Figure 7). Different criteria are applied to select the quadrat members of an isolated MRC, a few neighboring MRC, and the central MRC.

9. Limitations

One limitation comes from collecting and geocoding the restaurant data. First, although the data source is believed to be reliable for its large user base, about one third of the addresses were not found by the geocoder. Manual editing is very time-consuming for such a large dataset, and this task is likely to be automated by artificial intelligence. Second, it is difficult to verify how many restaurants were geocoded to the correct locations. Some of the not found addressed were edited to a nearby location. There is also uncertainties with the geocoder as 771 (2.24%) of the scraped data entries were still not found after editing, and 70 addresses were geocoded (0.2% of the geocoded) outside Beijing. Third, not all restaurants choose to provide information on Dianping. The dataset is merely a biased sample of all the restaurants in Beijing. In addition, web scraping can be both beneficial and harmful to the society (Din, 2015). In commercial practices, scraping data for profit is controversial and even illegal.

Another limitation comes from using artificially defined boundaries. To begin with, when constructing a fishnet for the quadrat count, the appropriate quadrat size is unknown and dependent on the point pattern. Large quadrats will generalize local variations while small quadrats will substantially increase the number of low values. Some quadrats on the edges were deleted after clipping, and this is dangerous for large quadrats. Based on the fishnet, the calculation of G*i* and Moran's I are subject to edge effects. In addition, these tools hypothesize that the attribute values follow a normal distribution. In the thesis, the Join_Count values of the quadrats obviously violate this assumption (Table 7).

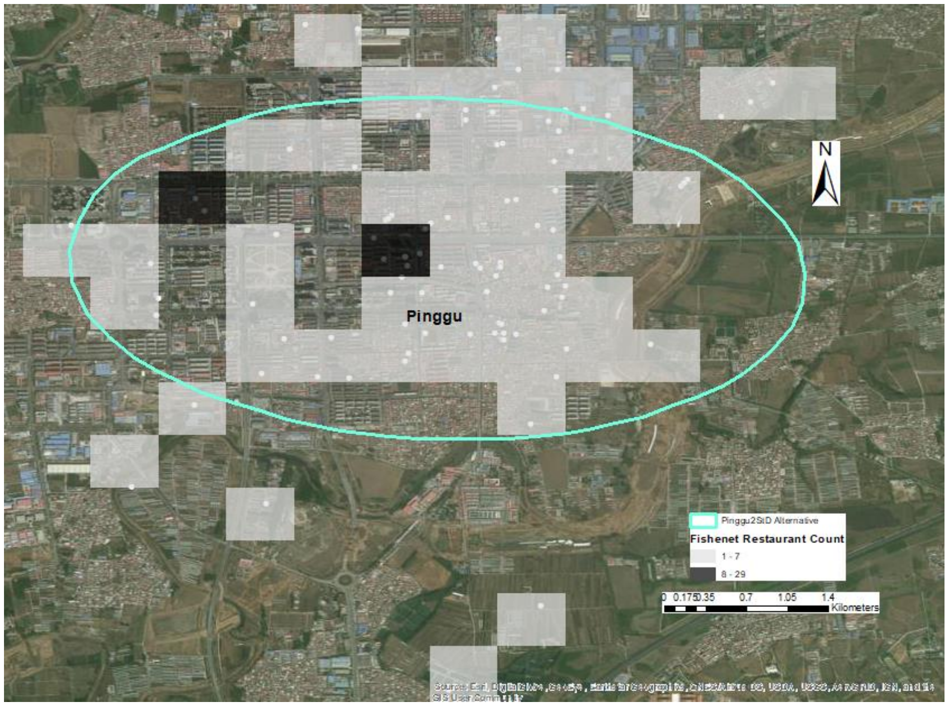


Figure 22 Alternative MCR Ellipse in Pinggu

Fundamental limitation lie in the method to describe a major restaurant cluster (MRC). First, there is a certain degree of subjectivity in the selection of MRC members. Second, the representation of an MRC is assumed to be an ellipse rather than irregular polygons, and the Directional Distribution tool does not allow inputs like 1.5 standard distance. Third, this approach is characterizing the quadrats rather than the restaurant points. Figure 12 shows an alternative MCR ellipse in Pinggu. The MRC members are the points within a group of populated quadrats. This ellipse accounts for the vertical distribution of the restaurants better than the result in Figure 15.

10. Application and Implications

While the concept of shangqu (trade area) implies the restaurant density across the study area, it turns out to be dependent on personal perception (Figure 4). Clustering of restaurants can be connected across shangqus and administrative boundaries.

The outlines of the MRCs will allow future researchers to study human activities within each of these significant areas and compare differences across space. One possible application is to present the data in a dashboard created by Pivot Table in Excel. Referring to Appendix B as an example, the user can choose the MRCs of interest, and view or compare activeness (percentage) of a type of cuisine in each area. The most popular cuisine types vary across the MRCs. While quick meal is welcomed everywhere, the demand for Beijing-style specialties is high in the Central MRCs. Japanese food is very profitable in this area. Korean food is prominent in the Wangjing MRC, while the overall market size of the Shunyi MRC is relatively small for investment. Within each MRC, future studies can also examine the innovation and evolution of each cuisine category over time.

References

ArcGIS. (2016 a). Directional Distribution (Standard Deviational Ellipse). Retrieved from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/directional-distribution.htm

ArcGIS. (2016 b). Hot Spot Analysis (Getis-Ord Gi*). Retrieved from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/hot-spot-analysis.htm

ArcGIS. (2016 c). Optimized Hot Spot Analysis. Retrieved from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/optimized-hot-spot-analysis.htm

ArcGIS. (2017). How Cluster and Outlier Analysis (Anselin Local Moran's I) works. Retrieved from http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-cluster-and-outlier-analysis-anselin-local-m.htm

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical analysis*, *27*(2), 93-115.

Beijing Municipal Bureau of Statistics. (2017). Beijing Statistical Yearbook. Retrieved from http://tjj.beijing.gov.cn/nj/main/2017-tjnj/zk/indexeh.htm

Beijing Municipal Bureau of Statistics. (2018). Annual data. Retrieved from http://www.bjstats.gov.cn/English/AD/

Craig, C. S. (1984). Models of the retail location process: a review. *journal of retailing*, *60*, 5-36.

Din, M. F. (2015). Breaching and Entering: When Data Scraping Should Be a Federal Computer Hacking Crime. *Brook. L. Rev.*, *81*, 405.

Fritz, C. E., Schuurman, N., Robertson, C., & Lear, S. (2013). A scoping review of spatial cluster analysis techniques for point-event data. *Geospatial health*, *7*(2), 183-198.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical analysis*, *24*(3), 189-206.

Ghosh, A., & MacLafferty, S. L. (1987). *Location strategies for retail and service firms*. Lexington Books.

Google Developers. (2018). Overview | Google Maps Geocoding API | Google Developers. Retrieved from https://developers.google.com/maps/documentation/geocoding/intro

Gould, P. (1970). Is statistix inferens the geographical name for a wild goose?. *Economic geography*, *46*(sup1), 439-448.

Litz, R. A., & Rajaguru, G. (2008). Does small store location matter? A test of three classic theories of retail location. *Journal of Small Business & Entrepreneurship*, *21*(4), 477-492.

Liu, W., & Xu, P. T. (2016). A Study on Influencing Factors of the Helpfulness of Online Reviews in O2O of Restaurant Industry--Based on Tobit Model. *Chinese Journal of Management Science*, *5*, 168-176.

Long, Y., & Thill, J. C. (2015). Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing. *Computers, Environment and Urban Systems*, *53*, 19-35.

Ma, W. (2016). *China's mobile economy: opportunities in the largest and fastest information consumption boom*. John Wiley & Sons.

Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: distributional issues and an application. *Geographical analysis*, *27*(4), 286-306.

O'sullivan, D., & Unwin, D. (2014). *Geographic information analysis*. John Wiley & Sons.

Pan, Y., Wu, D., & Olson, D. L. (2017). Online to offline (O2O) service recommendation method based on multi-dimensional similarity measurement. *Decision Support Systems*, *103*, 1-8.

Piore, M. J., & Sabel, C. F. (1986). The second industrial divide: possibilities for prosperity.

Porter, M. E. (2000). Location, competition, and economic development: Local clusters in a global economy. *Economic development quarterly*, *14*(1), 15-34.

Prayag, G., Landré, M., & Ryan, C. (2012). Restaurant location in Hamilton, New Zealand: clustering patterns from 1996 to 2008. *International Journal of Contemporary Hospitality Management*, *24*(3), 430-450.

Qi, W., Gao, Y., & Zhang, Q. (2017). Spatiotemporal Dynamics of Beijing's Urbanization Efficiency from 2005 to 2014. *Sustainability*, *9*(12), 2190.

Ramanathan, U., & Ramanathan, R. (2011). Guests' perceptions on factors influencing customer loyalty: An analysis for UK hotels. *International Journal of Contemporary Hospitality Management*, *23*(1), 7-25.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of applied probability*, *13*(2), 255-266.

Smith, S. L. (1985). Location patterns of urban restaurants. *Annals of Tourism Research*, *12*(4), 581-602.

Taneja, S. (1999). Technology Moves In. *Chain Store Age*, *75*(5), 136-137.

Wang, F. (2014). *Quantitative methods and socio-economic applications in GIS*. CRC Press.

Xiao, L., Mi, C., Zhang, Y., & Ma, J. (2017). Examining Consumers' Behavioral Intention in O2O Commerce from a Relational Perspective: an Exploratory Study. *Information Systems Frontiers*, 1-24.

Appendix A

The list of primary filter and secondary filters (the first two columns) for data collection was retrieved 1 Feb 2018 from http://www.dianping.com/beijing/ch10

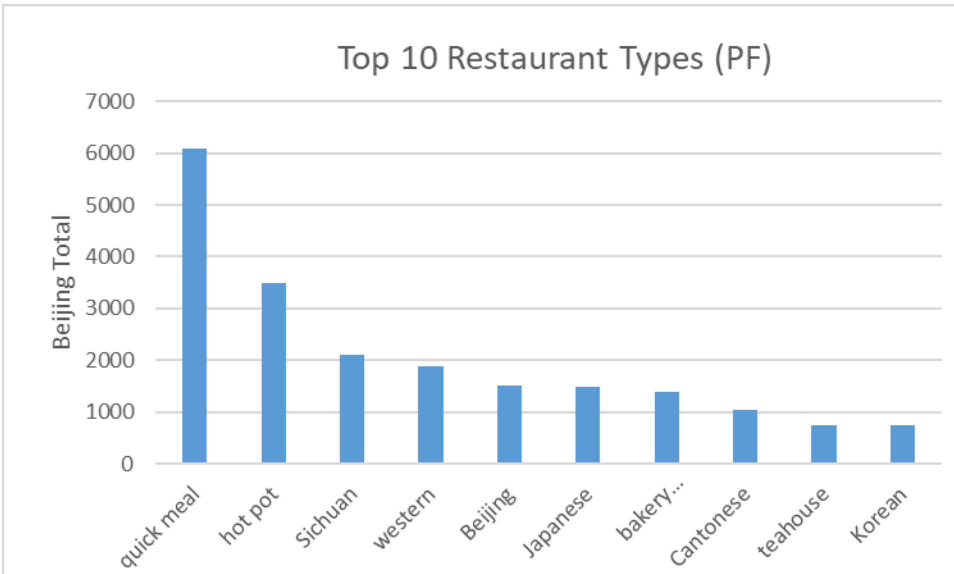
The other values were calculated in Excel 2016. Not found percentages before and after editing are highlighted for comparison. The table is ordered by primary filter ascending.

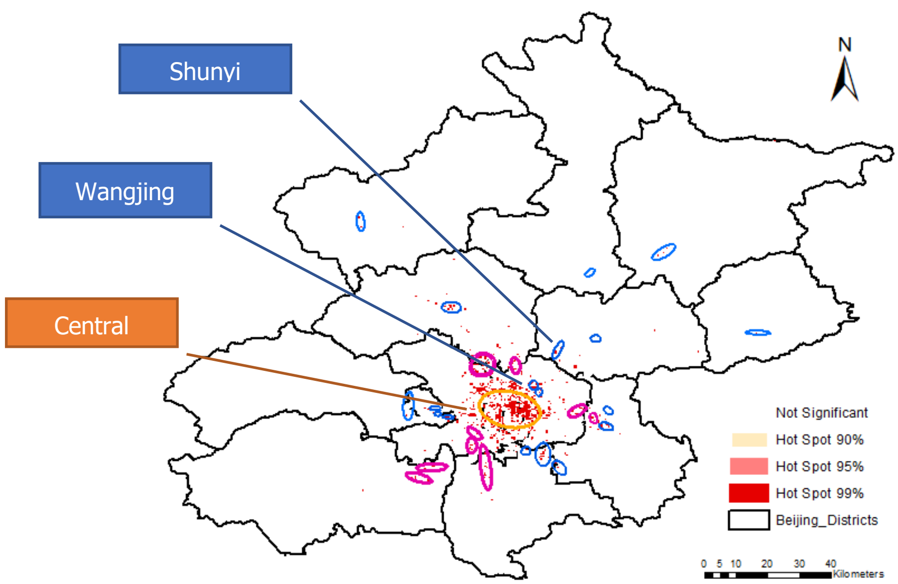
primary filter	secondary filter	total records (duplicates removed)	not found before editing	not found	not found after editing	final not found
sum		34485	11532	33.44%	771	2.24%
afternoon tea		750	145	19.33%	6	0.80%
Anhui		254	89	35.04%	9	3.54%
bakery and dessert	ice cream	700	227	32.43%	15	2.14%
bakery and dessert	drinks and desserts	741	189	25.51%	6	0.81%
barbecue		739	259	35.05%	19	2.57%
Beijing	farm style	750	379	50.53%	9	1.20%
Beijing	roasted duck	749	225	30.04%	17	2.27%
Beijing	nobility	71	8	11.27%	1	1.41%
buffet		737	199	27.00%	0	0.00%
Cantonese	dim sum	268	83	30.97%	2	0.75%
Cantonese	luxury	45	13	28.89%	2	4.44%
Cantonese	Chaoshan cuisine	104	29	27.88%	4	3.85%
Cantonese	Cantonese cuisine	692	162	23.41%	16	2.31%
coffehouse		723	221	30.57%	13	1.80%
crayfish		568	199	35.04%	13	2.29%
creative		431	132	30.63%	13	3.02%
Guizhou		91	25	27.47%	7	7.69%
home style		749	215	28.70%	22	2.94%
hot pot	fish	375	164	43.73%	12	3.20%
hot pot	small hot pot	500	198	39.60%	8	1.60%
hot pot	lamb spine	718	267	37.19%	16	2.23%
hot pot	skewers	574	211	36.76%	15	2.61%
hot pot	Chongqing	745	271	36.38%	24	3.22%
hot pot	instant-boiled mutton	749	262	34.98%	21	2.80%
Hubei		152	41	26.97%	1	0.66%
Hunan		743	246	33.11%	20	2.69%
Japanese	sushi	481	172	35.76%	12	2.49%
Japanese	barbecue	146	45	30.82%	1	0.68%
Japanese	quick meal	227	68	29.96%	1	0.44%
Japanese	Japanese cuisine	747	209	27.98%	5	0.67%
Japanese	noodles	114	29	25.44%	2	1.75%
Japanese	buffet	107	21	19.63%	0	0.00%

Jiangzhe		749	190	25.37%	11	1.47%
Korean		758	261	34.43%	13	1.72%
nonmeat		163	65	39.88%	2	1.23%
noodles		739	226	30.58%	6	0.81%
Northeastern Chinese		724	317	43.78%	30	4.14%
Northwestern Chinese		746	244	32.71%	17	2.28%
others	Shanxi cuisine	232	90	38.79%	1	0.43%
others	Jiangxi cuisine	125	49	39.20%	2	1.60%
others	other Chinese	750	245	32.67%	19	2.53%
others	Inner Mongolian cuisine	96	30	31.25%	3	3.13%
others	Mixed	185	55	29.73%	5	2.70%
others	teahouse	750	198	26.40%	5	0.67%
others	pub	749	155	20.69%	9	1.20%
others	Tibentan cuisine	15	0	0.00%	0	0.00%
private kitchen		476	190	39.92%	6	1.26%
quick meal	wonton	280	108	38.57%	2	0.71%
quick meal	malatang	747	287	38.42%	8	1.07%
quick meal	fried dumpling	253	97	38.34%	2	0.79%
quick meal	jiaozi	740	254	34.32%	24	3.24%
quick meal	fried sauce noodle	122	41	33.61%	4	3.28%
quick meal	congee	629	211	33.55%	19	3.02%
quick meal	delicatessen	738	246	33.33%	12	1.63%
quick meal	baozi	750	225	30.00%	16	2.13%
quick meal	fast food	745	213	28.59%	5	0.67%
quick meal	traditional Beijing	750	213	28.40%	11	1.47%
quick meal	snack	605	163	26.94%	17	2.81%
Russian		24	7	29.17%	0	0.00%
seafood		733	231	31.51%	10	1.36%
Shandong		271	82	30.26%	5	1.85%
Sichuan	grilled fish	750	275	36.67%	20	2.67%
Sichuan	dry pot	750	268	35.73%	20	2.67%
Sichuan	traditional	750	202	26.93%	16	2.13%
Southeast Asian		274	92	33.58%	4	1.46%
Taiwan		157	52	33.12%	4	2.55%
western	pizza	748	288	38.50%	14	1.87%
western	quick meal	750	247	32.93%	18	2.40%
western	steak	150	45	30.00%	3	2.00%
western	Latin barbecue	19	5	26.32%	0	0.00%
western	Italian	214	54	25.23%	2	0.93%
western	French	70	12	17.14%	1	1.43%
western	Middle Eastern cuisine	12	1	8.33%	1	8.33%
Xingjiang		648	207	31.94%	17	2.62%
Yunnan		289	88	30.45%	5	1.73%

Appendix B

A possible application using Excel Pivot Table: comparing restaurant information by MRC and by PF. Order by Count of Restaurants of each PF class.





MRC name	Central			
Restaurant Type	**Count of Restaurants**	**Percentage**	**Category Average Cost ¥**	**Total Reviews**
quick meal	**1813**	**17.54%**		**494253**
fast food	283	2.74%		146769
traditional Beijing	280	2.71%		147914
delicatessen	221	2.14%		20907
baozi	206	1.99%		30837
malatang	189	1.83%		32505
jiaozi	184	1.78%		32834
snack	166	1.61%		2800
congee	110	1.06%		47381
wonton	67	0.65%		7690
fried dumpling	55	0.53%		6484
fried sauce noodles	52	0.50%		18132
hot pot	**752**	**7.27%**		**491584**
instant-boiled mutton	199	1.93%		168180
Chongqing	156	1.51%		121911
skewers	150	1.45%		61993
lamb spine	113	1.09%		57871
small hot pot	88	0.85%		24283
fish	46	0.45%		57346
Japanese	**634**	**6.13%**	**181.23**	**537926**
Japanese cuisine	338	3.27%	199.03	391017
sushi	153	1.48%	184.03	66656
barbeque	59	0.57%	203.31	40965
noodles	52	0.50%	59.02	30383
quick meal	22	0.21%	33.92	3320
buffet	10	0.10%	226.00	5585

MRC name	Wangjing			

Restaurant Type	Count of Restaurants	Percentage	Category Average Cost ¥	Total Reviews
quick meal	**86**	**15.87%**		**14381**
delicatessen	13	2.40%		643
fast food	13	2.40%		5673
jiaozi	12	2.21%		1996
congee	10	1.85%		2074
malatang	10	1.85%		1048
traditional Beijing	10	1.85%		1954
baozi	7	1.29%		650
snack	6	1.11%		53
fried dumpling	4	0.74%		14
fried sauce noodles	1	0.18%		276
Korean	**82**	**15.13%**	**95.23**	**110861**
western	**41**	**7.56%**	**64.52**	**14720**
pizza	18	3.32%	73.13	8615
quick meal	15	2.77%	47.92	2954
Italian	7	1.29%	87.00	2530
French	1	0.18%	67.00	621
Sichuan	**32**	**5.90%**	**71.30**	**19921**
traditional	16	2.95%	72.88	11202
dry pot	11	2.03%	56.56	3154
fish	5	0.92%	92.80	5565
hot pot	**32**	**5.90%**		**19301**
Chongqing	14	2.58%		9093
instant-boiled mutton	8	1.48%		7355
small hot pot	4	0.74%		1034
skewers	3	0.55%		225
lamb spine	3	0.55%		1594

MRC name	Shunyi			

Restaurant Type	Count of Restaurants	Percentage	Category Average Cost ¥	Total Reviews
quick meal	**49**	**21.40%**		**4360**
snack	17	7.42%		204
delicatessen	7	3.06%		261
jiaozi	4	1.75%		1201
baozi	4	1.75%		544
congee	4	1.75%		198
fast food	4	1.75%		1012
fried dumpling	4	1.75%		639
malatang	2	0.87%		79
traditional Beijing	2	0.87%		222
wonton	1	0.44%		0
hot pot	**28**	**12.23%**		**5139**
skewers	10	4.37%		727
small hot pot	5	2.18%		483
Chongqing	5	2.18%		902
fish	4	1.75%		2498
lamb spine	2	0.87%		65
instant-boiled mutton	2	0.87%		464
western	**24**	**10.48%**	**66.31**	**6259**
pizza	11	4.80%	80.29	2234
quick meal	7	3.06%	40.33	3035
steak	4	1.75%	97.00	622
Italian	1	0.44%	63.00	368
French	1	0.44%		0
Sichuan	**22**	**9.61%**	**71.85**	**11951**
dry pot	10	4.37%	57.63	3305
fish	9	3.93%	91.67	7917