Keep Your Enemies (Competitors) Closer:

A study of competition clustering in the hospitality industry in urban areas.

Zewen Feng

Introduction:

The proverb "keep your friends close but keep your enemies closer" imparts wisdom on not only the military strategists in the ancient times, but also economists. One would consider it optimal for restaurants to be evenly distributed across urban population centres to reduce competition among businesses and servicing a wider segment of population. However, restaurants competing in the same culinary category tend to cluster and their choices of location display limited spatial dispersion.

Clustering implicitly helps restaurants collectively advertise their cuisine and achieve a shared marketing objective. For instance, one might automatically think of going to the Chinatown for Chinese food without searching for Chinese restaurants in their own neighbourhood. However, in an informationally efficient environment where one can easily search for the closest point-of-interest (POI), spatial proximity incurs more costs for customers in the form of longer average travelling time. Thus, it is important to examine whether clustering is still optimal in the digital age.

This report documents the clustering effect of restaurants, cafés and bars in London. It further sheds some light on whether opening restaurants away from competitors is economically sensible. The findings of this report could be useful in optimising urban zoning policies to minimise travelling times and improve ease-of-access by the population.

Data Procurement:

Data Source

Locational data provided by Foursquare API can help identify the locations of points-of-interest in major cities. In addition to the location of a point-of-interest (described by its coordinates), the dataset also identifies its type, user reviews, and customer ratings. Foursquare allows developers to send requests to extract information about points of interest near a given location. We specify the central London as the starting point for our search queries. Our search queries include three types of point-of-interest: coffee, food, and drinks.

Due to the focused scope of this study on urban population centres and limited access to API granted by the data provider, only restaurants within the 2km radius of a London's CBD will be examined. The queries returned 1,224 points-of-interest meeting the criteria. Figure 1 visualises these POIs on a map of London. The red, yellow, and blue marks represent coffee shops, restaurants (or food vendors), and bars, respectively.

Data Cleaning

The data provided are compiled into json files. We parse the json data and keep the following attributes for all entries obtained:

- Venue ID: the unique alpha-numerical ID for the venue

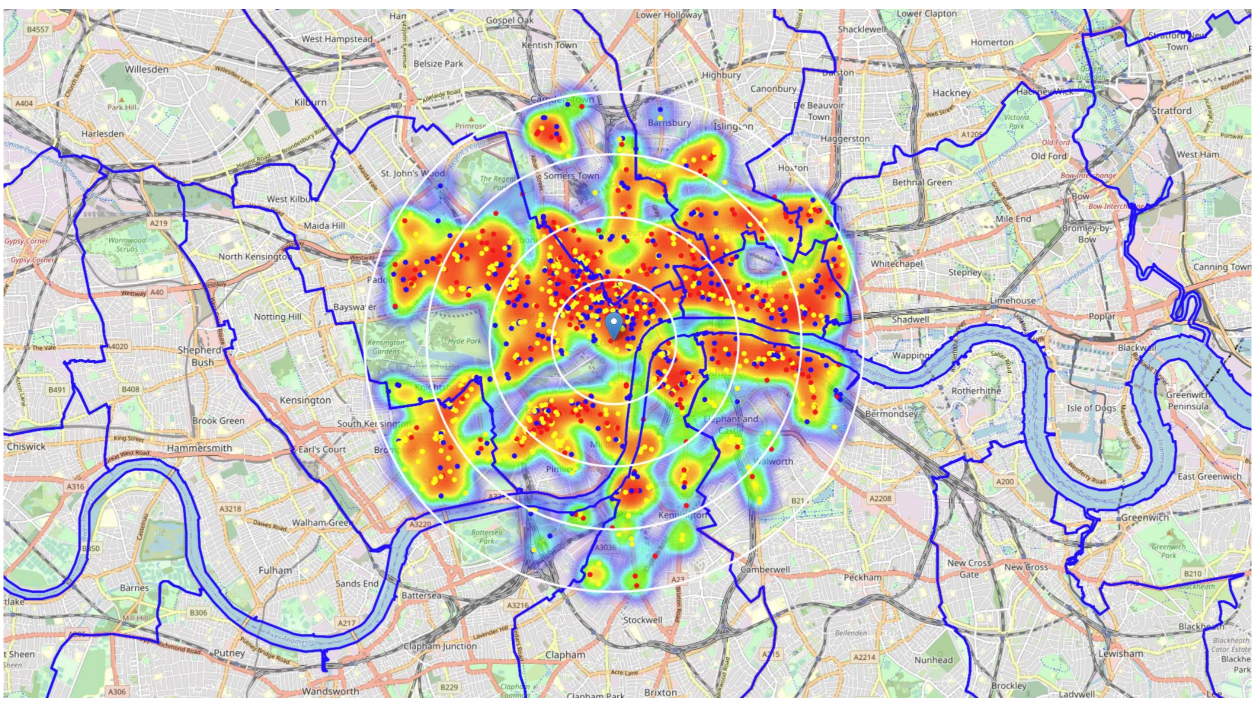
- Geographic coordinates specified by the venue's latitude and longitude
- The name of the POI
- The category of POI, including coffee shops, restaurants, and bars
- Customer rating of the POI, on a scale of one to ten

The queries returned numerous entries that are irrelevant to this study. For instance, the obtained dataset contains information about nearby museums, movie theatres, and shopping centres. We filter these venues out by only including POIs with their category name containing the words 'coffee', 'café', 'restaurant', and 'bar'. The filtered sample contains 1,224 POIs, including 304 coffee shops, 353 bars, and 567 restaurants.

Exploratory Data Analysis

Figure 1 presents a map of London showing the POIs identified with 4km radius from central London. The blue borders outline the borders between boroughs and the river Thames. The centre is located on the Trafalgar Square, within the borders of the city of Westminster. Each dot represents the location of one POI, with the colours red, blue, and yellow representing coffee shops, bars, and restaurants, respectively. A heatmap is projected on the map, warmer colours mean that POIs are more densely situated. Clearly, the central London on the northern bank of the river houses more POIs than the southern bank.

Figure 1 Map of London with Points-of-Interest



Customer rating is an important metric that measures the quality of the service/product provided by a business. Such metric is ever more essential to the success of a business in the digital age where users could easily filter through hundreds of alternatives based on the ratings. For this reason we consider the customer rating a close proxy to measure the success and popularity of a POI.

User rating information is summarised in Figure 2. Restaurants receive higher ratings than bars and coffee shops, a one-way ANOVA reveals that the average rating of 7.99 for restaurants is 0.27 and 0.42 higher than those for bars and coffee shops, respectively. These differences in average ratings among groups are significant at 5%. Figures 3 and 4 plot the ratings of POIs and their distances from the centre of London and centre of the clusters, respectively. There is little indication of whether linear relationship exists between rating and distances.

Figure 2 Customer Ratings

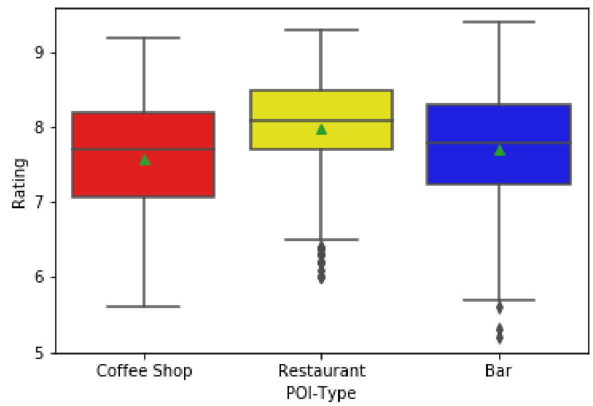


Figure 3 POI Rating and Distance from Centre of London

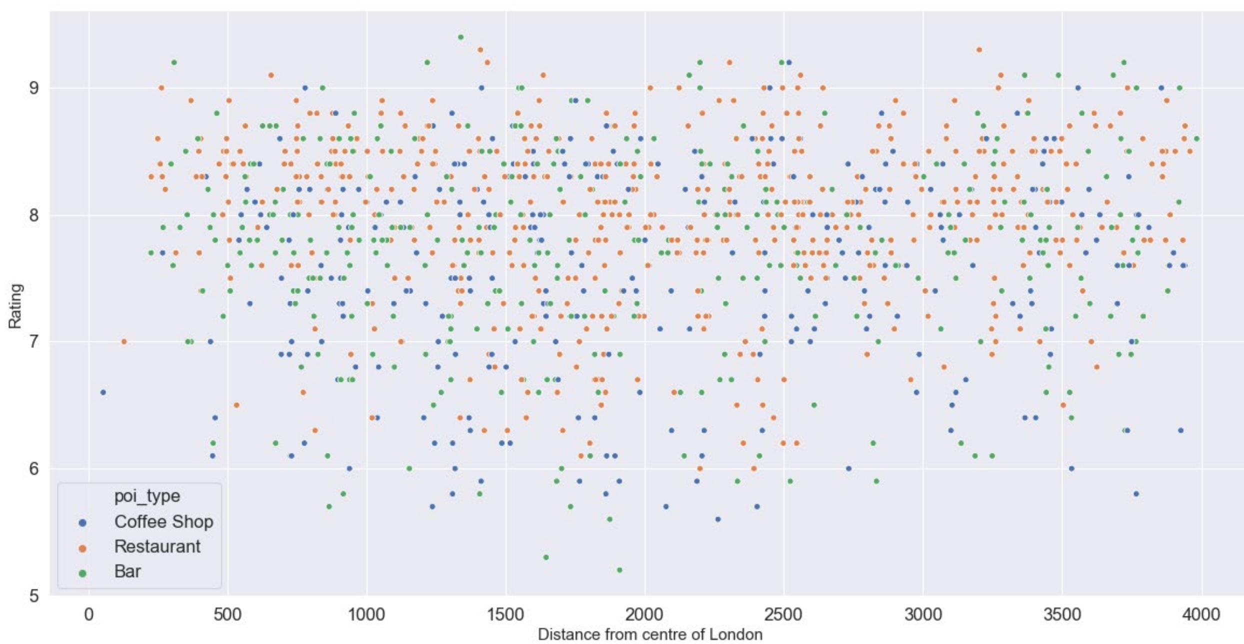
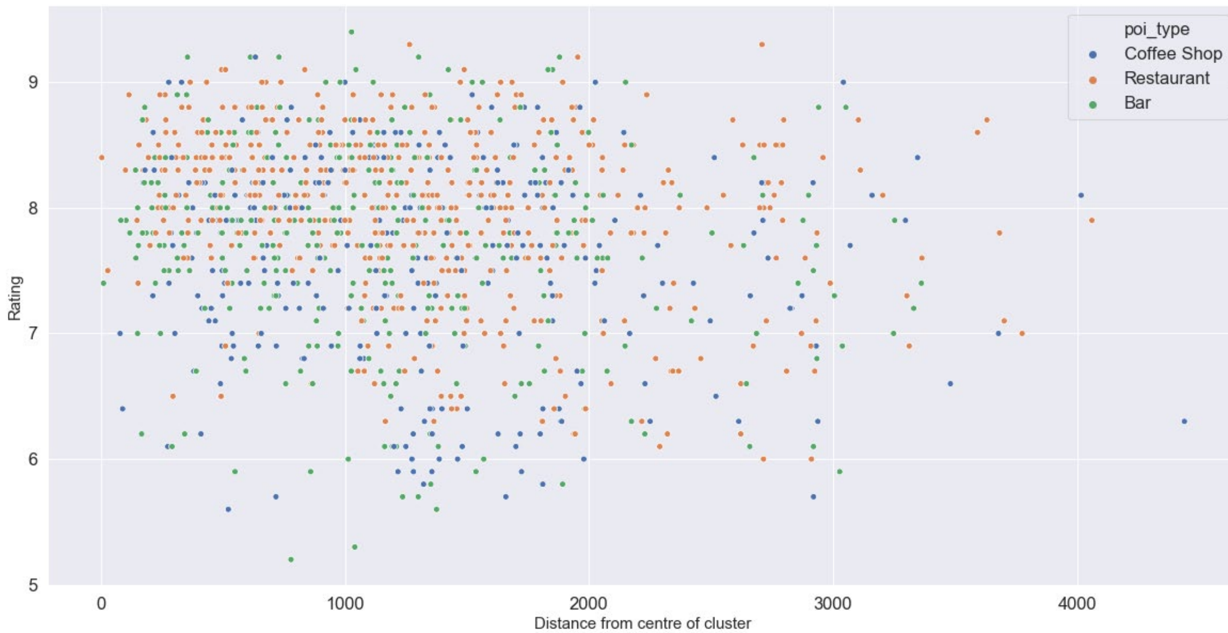


Figure 4 POI Rating and Distance from Centre of Cluster



Methodology

We first utilise the K-Means algorithm to separate these POIs into 6 clusters and compute the coordinates of the centroids of these clusters. For a POI i assigned to cluster k, we compute its distance $distCentroid_i$ from the centroid C_k using the Vincenty's formulae.

We move on to address the key issue identified: whether there is a relationship between a POI's success and its relative proximity to its competitors. This issue could be interpreted and examined from various aspects. Admittedly, a POI's rating could well be affected by numerous endogenous factors, such as the quality and novelty of the products and services served, the experiences and skills of the employees, and the prices of the products and services. Unfortunately, we do not have data on such endogenous variables that are idiosyncratic to POIs. Thus, the scope of this study is inherently limited to how the location of a POI is related to its rating.

We first examine whether a POI's rating is affected by its distance from the centre of the city. This is easily achieved through an OLS analysis. Since it is already recognised that different types of POIs have significantly different ratings, we include a vector of dummy variables $TypeControl$ to control for the inter-group rating differences.

$$OLS1\text{: } Rating_i = \beta_0 + \beta_1 Distance\_to\_centre_i + \beta TypeControls_i + \epsilon_i$$

Normally one would expect the rents for locations near the city centre to be high, thus it is likely that only more successful businesses with high customer ratings are able to survive close to the city. However, since our dataset only covers a radius of 4kms from the centre of London, a relatively small radius considering the size and global prestige of London, rents faced by the POIs in our dataset are all considerably high.

For this reason, we divide the sample into two groups based on whether they are located within 2kms of the centre. The dummy variable far_2km takes the value of 1 if the POI is more than 2kms away from the centre, and 0 otherwise. This variable is then interacted with the distance variable previously defined.

$$OLS2\text{:}\ Rating_i = \beta_0 + \beta_1 Distance\_to\_centre_i\ + \beta_2 far\_2km + \beta_3 far\_2km \times Distance\_to\_centre + \beta TypeControls_i + \epsilon_i$$

In a separate specification, we examine whether POIs located away from its competitors of the same type will have higher ratings. We count the number of competitors that are located less than 100 meters from a POI and run the following regression. Average distance to these competitors are also computed.

$$OLS3\text{:}\ Rating_i = \beta_0 + \beta_1 NumOfComp_i\ + \beta TypeControls_i + \epsilon_i$$

$$OLS4\text{:}\ Rating_i = \beta_0 + \beta_1 \boldsymbol{DistSameType}_i\ + \beta TypeControls_i + \epsilon_i$$

Empirical Results:

OLS results are summaries in Table 1. The first specification of OLS indicates that a POI's distance from the centre of London has no significant impact on its rating. This is potentially because that most POIs are located close to the centre and there is little amount of variation of distance for the variable to be significant.

The results for specification 2 shows that distant POIs receive higher ratings the more distant they are allocated from the centre. According to the principle of marginality, the coefficient for the interaction term suggests that on average, POIs located more than 2 kms from the centre of London have 0.3611 higher rating for each additional kilometre of distance.

The last OLS specification yields some interesting results. More competitors within a short range decreases the rating of a POI. This result suggests that businesses should avoid locating too close to its competitors.

Overall, the results from the above OLS combined indicate that for POIs to receive higher ratings, they need to locate more distantly from the centre of the city, and away from its competitors.

Table 1 OLS Results

		$\boldsymbol{Rating}$	
$\boldsymbol{Distance\_to\_centre}$	0.0177	-0.2250***	
$\boldsymbol{far\_2km}$		-0.6148***	
$\boldsymbol{far\_2km}$ $\times \boldsymbol{Distance\_to\_centre}$		0.3611***	

$\boldsymbol{NumOfComp}$				-0.0200*
$\boldsymbol{DistSameType}$			1.0441	
$\boldsymbol{TypeControls(Restaurant)}$	0.2719***	0.2851**	0.2361***	0.2790*
$\boldsymbol{TypeControls(Coffee)}$	-0.1454***	-0.1340***	-0.1787**	-0.1350**

Conclusion

This report examines a pertinent issue on urban planning and strategic choice of locations for POIs. We investigate the statistical relationship between the physical proximity of POIs with each other and their customer ratings. Conventional wisdom suggests that businesses offering similar products and services (such as coffee, food, and alcoholic drinks) with limited viability for differentiation should be located away from each other. However, the heatmap in this report clearly indicates that businesses instead cluster around several hotspots in the city centre. One argument for clustering is that it helps businesses competing in the same category boost their collective presence, effectively resulting in free marking. The downside of this clustering is apparent: businesses must closely compete against each other to acquire market shares.

The advances in digital technologies cause a paradigm shift, which necessitates a revision to the decision by businesses to cluster. Digital presence drastically lowers the cost of marketing, inherently reducing the effectiveness of sacrificing market shares for market presence in urban centres. We thus investigate whether it is still more optimal for businesses to cluster around their competitors.

We find that businesses that are away from the city centre tend to enjoy higher customer ratings. In addition, ratings are even higher when the businesses have fewer competitors in close proximity. Our finding directly challenges the notion that businesses should cluster in urban centres. However, we acknowledge that there might be some unobserved endogenous factors not reflected by user ratings that could affect a business's success. A possible extension could be made to include more business-specific factors in order to boost robustness of our study.