Are central-zone restaurants better for consumers? -An analytical approach

Conference Paper · December 2021

DOI: 10.1109/BigData52589.2021.9671330

5 authors, including:

Filip Markoski
Ss. Cyril and Methodius University in Skopje
5 PUBLICATIONS **8** CITATIONS

Lasko Basnarkov
Ss. Cyril and Methodius University in Skopje
57 PUBLICATIONS **322** CITATIONS

Biljana Risteska Stojkoska
Ss. Cyril and Methodius University in Skopje
75 PUBLICATIONS **1,535** CITATIONS

Petre Lameski
Ss. Cyril and Methodius University in Skopje
103 PUBLICATIONS **957** CITATIONS

Are central-zone restaurants better for consumers? - An analytical approach

Filip Markoski, Lasko Basnarkov, Biljana Stojkoska, Petre Lameski, Eftim Zdravevski
Ss. Cyril and Methodius University in Skopje, Faculty of Computer Science and Engineering
Skopje, Macedonia
filip.markoski45@gmail.com, lasko.basnarkov@finki.ukim.mk, biljana.stojkoska@finki.ukim.mk,
petre.lameski@finki.ukim.mk, eftim.zdravevski@finki.ukim.mk

***Abstract*—Analysis of restaurant location has significant value both for owners, and for consumers. In this paper we examine the statistical relationships of restaurant properties and qualities in order to establish their relevance for the success of a restaurant. For this purpose, we use information-based methods and a dataset obtained from the TripAdvisor platform containing data about restaurant facilities situated in London. The dataset is rich with numerous qualitative features describing the restaurants. Through the use of statistical and geographical methods we explore the usefulness of geographic visualisation of restaurant qualities, and examine the statistical relationships among restaurant descriptors. Our results demonstrate the effectiveness of information methods for researching economic and non-economic phenomena.**

***Index Terms*—London, Restaurants, Centralism, Information methods**

I. Introduction

All restaurants possess a certain collection of qualities which implore the question of which are the most relevant factors to a restaurant's success. Understanding and projecting the restaurantscape is important, and some exiting articles analyse the influence of neighbourhood socio-demographic characteristics on restaurant location [1].

Knowing which properties of the restaurant are causally related to the economic performance of the restaurant could help owners and decision makers make the right choices and focus on the right properties in order to achieve success. In [2] authors analysed why restaurants fail, particularly focusing on the affiliation, location, and size on restaurant failures. Therefore, knowing which locations would be most profitable is quite important, as analysed in [3], where multi-criteria selection for a restaurant location in Taipei was performed. Similarly, [4] investigates the clustering patterns of restaurant locations.

In this paper we aim to further extend the body of knowledge by providing an exploration into the various properties of the restaurants and compare their attributes using statistical analysis approaches to discover to what degree a certain restaurant aspect is relevant to the performance of a given restaurant.

An additional purpose of this research is to demonstrate how effective information methods can be in the process of researching problems in some economic and non-economic activities as a cheaper and faster alternative, which is in line with the narrative discussed in [5].

Our work resembles the work of [6] in which the authors explore whether the performance-related patterns of a given restaurant vary with different regions. Additionally, previous efforts have been made to explore some of the cultural aspects of a given restaurant facility [7]. The authors of [8] explored restaurant ownership turnover rates to discover that there are marginal differences in restaurant failures between franchise chains and independent restaurants. Similarly to the work of [9], we incorporate restaurant customer expectations and perceptions and attempt to discover a pattern between them and the performance of the restaurant. The authors of [10] examines the re-location of upper-class and middle-class schools from central London to the suburbs which we deem as relevant to our work due to the postulated relationship between wealth and luxury. Another study presented in [11] note that there are also internal factors such as overconfidence of the owners or unfitness to lead the restaurant that could be significant indicator for the success of a restaurant.

The main contributions of this paper are: the exploration of the usefulness of geographic visualisation of restaurant qualities via the use of thematic maps, the examination of statistical relationships among restaurant descriptors through the use of correlation, associations, and other analyses and finally, demonstration of the effectiveness of information methods for researching economic and non-economic activities, as a cheaper and faster alternative to the conventional survey approaches.

II. Methods

A. Data Set

From the data collected approximately in the span of March 2021 from the TripAdvisor platform, we were able to identify 19,850 restaurants situated in London, which is in accordance with the number of restaurants reported in [12]. The data, which was collected during March 2021, contains various restaurant features such as restaurant price range, restaurant address, restaurant rating, as well as, individual restaurant reviews accompanied with a restaurant review rating. We found that only 12,410 contained enough reviews to obtain an accurate rating of the restaurant, calculated as an average of the reviews submitted to the TripAdvisor platform. The currently explored dataset was chosen for practical reasons as it is related to London, a world metropolis. If we were to extend the dataset with data from other locations, big data architectures would become necessary for efficient and timely processing of it [13]. In turn, it would entail the use of efficient algorithms for cluster-size and cost optimization [14], [15]. As these kinds of datasets contain often multi-modal data comprised of time series, nominal, ordinal and categorical attributes, appropriate preprocessing methods, such as the weight of evidence, are required to transform them into numeric data [16]. The feature selection on such big datasets should also consider the use of scalable algorithms [17].

B. Feature extraction

From the available data we extract the information about which London Borough the restaurant belongs to by using the restaurant address obtained from TripAdvisor. We use the Nominatim's reverse geocoding API [18], which uses Open Street Maps [19] to obtain the address given the coordinates of a location. By using the restaurant price range found in the restaurant data source we created a categorical feature which contains the following set of unique values: {'High-priced', 'Low-priced', 'Middle-priced', 'Unknown'}. Within our restaurant data source 23 unique type categories for restaurants can be found such as 'Bars', 'Cafe', etc. On the other hand, cultural categories are nominal values which represent the cultural, ethnic or geographic attribute that a restaurant may poses (ex. 'British', 'Asian' etc.) There are 71 unique cultural categories. For each unique nominal value of the categories we create a separate binary feature which denotes whether the current restaurant poses the property or it does not. Thus, 94 binary features were created for the restaurant dataset.

III. Results

In this section we present the results obtained from the application of the methods upon the acquired and pre-processed restaurant dataset. The results are thus formed from our analyses using geographical visualisation techniques and statistical relationship approaches.

A. Restaurant Rating Distributions

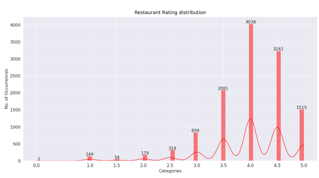

Fig. 1. Restaurant Rating Distribution

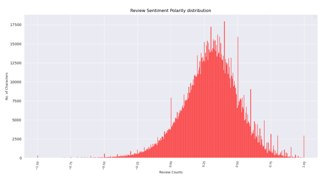

Fig. 2. Review Sentiment Polarity Distribution

In Figure 1. we can examine that most of the restaurants have a restaurant rating greater or equal to 3.5 and in Figure 2, we depict the distribution of the sentiment polarity of each textual restaurant review submitted. The sentiment polarity is an individual coefficient per restaurant review whose value is between -1 and $+1$, with the former denoting negative sentiment and the latter denoting positive sentiment.

B. Geographical Visualisations

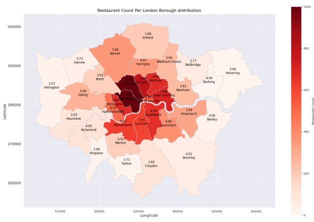

Fig. 3. Cartodiagram Restaurant Count Per London Borough Distribution

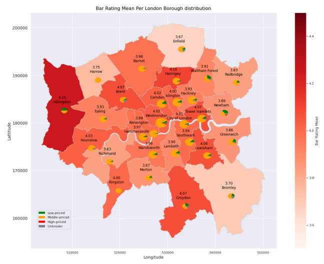

Fig. 4. Cartodiagram Bar Rating Mean Per London Borough Distribution

In Figure 3. we present the number of restaurants for each of the London boroughs. It can be seen that the number of restaurant facilities is densest in the central areas and decreases towards the periphery. Furthermore, judging from Figure 4. bars with the highest price ranges are predominantly located in the centre of London whilst cheaper alternatives are more present in the peripheral areas. It can also be seen that objects with medium price ranges, prevail in the central zones, but are located on a significantly larger territory than the central parts of the agglomeration. Also, restaurant facilities with lower prices are located throughout the territory, but their number in the central zones of agglomeration is smaller.

C. Correlation Analysis

The results obtained by applying the Pearson's correlation coefficient are as follows: The top most correlated features with the 'RestaurantRating' feature in descending order are: 'RestaurantPriceRange' (-0.1357), 'Chinese' (-0.1256), 'Mediterranean' (0.0964), 'Cafe' (0.0879), 'European' (0.0687). The top most correlated features with the 'RestaurantPriceRange' feature in descending order are: 'RestaurantRating' (-0.1357), 'Asian' (-0.1005), 'Cafe' (-0.0989), 'Bar' (-0.0954), 'British' (-0.0951). The top most correlated features with the 'LondonBorough' feature in descending order are: 'Indian' (0.0599), 'European' (-0.0394), 'Japanese' (-0.0389), 'Persian' (-0.0376), 'Bangladeshi' (0.0353).

D. Associations Analysis

1) Cramér's V

By applying the Cramér's V method upon the features of the restaurant dataset, we calculated that the top most associated features with the 'RestaurantRating' feature in descending order are: 'UnknownCulture' (0.1668), 'Bar' (0.1584), 'Chinese' (0.1421), 'British' (0.1240), 'Mediterranean' (0.1120), 'European' (0.1084). The top most associated features with the 'RestaurantPriceRange' feature in descending order are: 'Fast Food' (0.2953), 'Cafe' (0.2344), 'UnknownCulture' (0.2103), 'Bar' (0.1896), 'European' (0.1706), 'UnknownType' (0.1365), 'RestaurantRating' (0.1177), 'French'

(0.1163). The top most associated features with the 'LondonBorough' feature in descending order are: 'RestaurantPriceRange' (0.1628), 'Turkish' (0.1474), 'Indian' (0.1363), 'Pakistani' (0.1278), 'Beer restaurants' (0.1273), 'European' (0.1239), 'UnknownCulture' (0.1072), 'Bar' (0.1043).

2) Theil's U

According to the measurements results obtained by applying the Theil's U method upon the features of the restaurant dataset, the top most associated features with the 'RestaurantRating' feature calculated using the Theil's U measure in descending order are: 'Salvadoran' (0.2584), 'Coffee & Tea' (0.2161), 'Dining bars' (0.2017), 'Nigerian' (0.1713), 'Campania' (0.1288). The top most associated features with the 'RestaurantPriceRange' feature calculated using the Theil's U measure in descending order are: 'Ecuadorean' (0.1551), 'Fast Food' (0.1375), 'Egyptian' (0.1071). The top most associated features with the 'LondonBorough' feature calculated using the Theil's U measure in descending order are: 'Beer restaurants' (0.5242), 'Salvadoran' (0.3007), 'Dining bars' (0.2900), 'Nigerian' (0.2808), 'Chilean' (0.2785).

3) Correlation Ratio

The correlation ratio between the numerical 'RestaurantRating' and categorical 'LondonBorough' features equals 0.09 with a two-tailed p-value of 0.276, whilst the correlation ratio between the numerical 'RestaurantRating' and categorical 'RestaurantPriceRange' features equals 0.049 with a two-tailed p-value of almost 0.000. The calculation of the p-value relies on the assumption that each dataset is normally distributed.

IV. Conclusion and Discussion

Based on the regional and spatial distribution, the functional aspects of the city, the population density of the individual settlements and neighbourhoods and other aspects included in this research, it can be concluded that:

- The attendance and cost of the facilities is higher in the central parts of the city, whilst it is decreasing towards the periphery. This density within the central area is most likely due to influx of tourists which prioritise visiting monuments typically located in the heart of London. Such are administrative political buildings, business located offices and headquarters, cultural buildings, traffic jams and more.
- We may extrapolate that there are fewer tourist visits to the periphery, more locals, more commercial buildings, sports facilities, ecological aspects, etc. as the locals in the periphery are more family-oriented and follow a suburban lifestyle, which is in accordance with the findings of the [10].
- There are various restaurant facilities based on the aspects of restaurant type and cultural affiliation.

The main drawback of the study is that internal factors of the restaurant such as staff training, background and owners or managers capabilities etc., were not included in the data due to their unavailability. According to some studies, these factors could also be important for estimating the success of a restaurant.

Acknowledgements

The work presented in this paper was partially funded by the University of Ss. Cyril and Methodius in Skopje, Faculty of Computer Science and Engineering.

References

[1] Y. Yang, W. S. Roehl, and J.-H. Huang, "Understanding and projecting the restaurantscape: the influence of neighborhood sociodemographic characteristics on restaurant location," *International Journal of Hospitality Management*, vol. 67, pp. 33–45, 2017.

[2] H. Parsa, J. Self, S. Sydnor-Busso, and H. J. Yoon, "Why restaurants fail? part ii-the impact of affiliation, location, and size on restaurant failures: Results from a survival analysis," *Journal of Foodservice Business Research*, vol. 14, no. 4, pp. 360–379, 2011.

[3] G.-H. Tzeng, M.-H. Teng, J.-J. Chen, and S. Opricovic, "Multicriteria selection for a restaurant location in taipei," *International journal of hospitality management*, vol. 21, no. 2, pp. 171–187, 2002.

[4] G. Prayag, M. Landré, and C. Ryan, "Restaurant location in hamilton, new zealand: clustering patterns from 1996 to 2008," *International Journal of Contemporary Hospitality Management*, 2012.

[5] D. Sui and R. Morrill, "Computers and geography: From automated geography to digital earth," in *Geography and technology*. Springer, 2004, pp. 81–108.

[6] J. Wang and X. Yan, "How location matters in restaurant success?" in *PACIS*, 2017.

[7] N. O. Madichie, "Nigerian restaurants in london: bridging the experiential perception/expectation gap," *International Journal of Business and Globalisation*, vol. 1, no. 2, pp. 258–271, 2007.

[8] H. Parsa, J. T. Self, D. Njite, and T. King, "Why restaurants fail," *Cornell Hotel and Restaurant Administration Quarterly*, vol. 46, no. 3, pp. 304–322, 2005.

[9] O. Mhlanga, "Measuring restaurant service quality in east london, south africa: a comparison of restaurant customer expectations and perceptions." *African Journal of Hospitality, Tourism and Leisure*, 2018.

[10] S. Gamsu, "Moving up and moving out: The re-location of elite and middle-class schools from central london to the suburbs," *Urban Studies*, vol. 53, no. 14, pp. 2921–2938, 2016.

[11] A. A. Camillo, D. J. Connolly, and W. G. Kim, "Success and failure in northern california: Critical success factors for independent restaurants," *Cornell Hospitality Quarterly*, vol. 49, no. 4, pp. 364–380, 2008. [Online]. Available: https://doi.org/10.1177/1938965508317712

[12] Public. Number of restaurants. [Online]. Available: http://www.worldcitiescultureforum.com/data/number-of-restaurants

[13] E. Zdravevski, P. Lameski, C. Apanowicz, and D. Slezak, "From big data to business analytics: The case study of churn prediction," *Applied Soft Computing*, vol. 90, p. 106164, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494620301046

[14] M. Grzegorowski, E. Zdravevski, A. Janusz, P. Lameski, C. Apanowicz, and D. Slezak, "Cost optimization for big data workloads based on dynamic scheduling and cluster-size tuning," *Big Data Research*, vol. 25, p. 100203, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214579621000204

[15] E. Zdravevski, P. Lameski, A. Dimitrievski, M. Grzegorowski, and C. Apanowicz, "Cluster-size optimization within a cloud-based etl framework for big data," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 3754–3763.

[16] E. Zdravevski, P. Lameski, A. Kulakov, and S. Kalajdziski, "Transformation of nominal features into numeric in supervised multi-class problems based on the weight of evidence parameter," in *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2015, pp. 169–179.

[17] E. Zdravevski, P. Lameski, A. Kulakov, S. Filiposka, D. Trajanov, and B. Jakimovski, "Parallel computation of information gain using hadoop and mapreduce," in *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2015, pp. 181–192.

[18] Nominatim. [Online]. Available: https://nominatim.org/

[19] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org ," https://www.openstreetmap.org, 2017.