A GIS AND SPATIAL-ECONOMETRIC STUDY OF CTRIP-LISTED HOTELS IN SAMARKAND HERITAGE PROXIMITY, PLATFORM POPULARITY, AND SPATIAL SPILLOVERS

ZHAO YING
'Silk Road' International University of Tourism and Cultural Heritage, Samarkand, Uzbekistan
Tianjin College of Commerce, Tianjin,China
ORCID 0009-0009-2530-1823
https://doi.org/10.5281/zenodo.22111736

Abstract. *Our investigation examines the hotel inventory and platform popularity in Samarkand based on the Ctrip data of August 15th, 2026. The data consist of 323 listings, among which 302 have ratings and 4,661 comments. The hotel names are chosen from OpenStreetMap/Photon and after a thorough check, 171 different places are obtained. The spatial distribution is clustered (nearest-neighbour index 0.734, p<0.001); the peak of kernel-density is situated 0.12 km away from Registan Square. No spatial autocorrelation exists in the hotel scores (Moran's I=-0.005, p=0.882). The logarithm of the number of reviews shows a positive spatial relationship (rho=0.43, p=0.003); there is a positive correlation between the hotel score and the platform attention. Cleanliness and service of the rooms are referred to more frequently than the words associated with historical sites. The results can differentiate the physical concentration from the effect of platform-mediated popularity, offering a clear basis for the East Asia policy of Ctrip.*
Keywords: *Ctrip; hotel popularity; GIS; online reviews; Samarkand*

Introduction

Online travel agencies do not sell only hotel rooms. They also assess the conditions of the properties, keep the record of the comments, translate the information about the destination and make some hotels more popular than others. For a historic place, this online work is related to its real geographical position and the transportation facilities. A hotel may be close to the historic site but does not receive much online attention, whereas another property situated in the suburban area can be noticed because of its reputation, convenient transportation services or favorable comments. Consequently, there exists a link between the hotel's position and its popularity on the website, but these two situations are differently considered from an analytical viewpoint.

Samarkand possesses a special atmosphere for this reason. The Registan Square is an important part of the historical tourist industry, and besides, the railway route and the Silk Road Samarkand resort area are also attractive places.

The paper puts forward three questions. Firstly, what is the concentration degree of Ctrip-listed hotels around the heritage core? Secondly, is there any spatial dependence in the review volume after controlling for the observable accessibility and quality factors? Thirdly, which service aspects and guest-experience topics characterize the platform's review content? The aim is to integrate point-pattern analysis, spatial econometrics and multilingual review information into one specific design.

Conceptual background

The location of the hotel depends on its nearness to the tourist places, transportation stations, commercial districts, land market and the planning regulations. In the investigation of urban tourism, the distribution of attractions is usually located along the corridors instead of being dispersed at random. In the ancient cities, the easy accessibility to the monuments decreases the travelling expenses for the tourists and improves the attractiveness of walking, but the congested traffic and limited land may lead to the building of big structures in the outskirts. Hence, Samarkand follows two main rules: a small-scale supply close to the Registan and a planned resort area in the northeast.

Nearest-neighbour and kernel-density statistics characterize this morphology, but they cannot clarify the platform performance. The review count is a total visibility indicator which is influenced by the age of the hotel on the platform, its capacity, demand and ranking exposure. The nearby hotels may have the same micro-markets in the destination, the same transport routes and the same search-page visibility, resulting in a spatially associated popularity process although their ratings are different.

Ctrip is important because its Samarkand page combines Chinese hotel information together with reviews in various scripts. Some previous studies about Chinese online travel agencies have suggested that the importance of topics and semantic relations differs according to the website. Besides, the comparative investigations indicate that the user generated-content cannot be applied directly to other sites. Hence, the language of the reviews and the service themes are regarded as the characteristics of Ctrip's data rather than describing all the tourists to Samarkand.

Data and methods

The hotel sheet contains 323 entries concerning Samarkand. The ratings of 302 sites are recorded. The number of reviews is given, but the distribution is greatly unequally: the median hotel has received 2 reviews and the average is 16.67. Each of the four hotel-related sub-scores, like location, facilities, service and room, is evaluated on a five-point scale. The review sheet comprises 4,661 items including the hotel ID, travel group, stay month, publication time, score, text, helpfulness, photographs and lacking original metadata.

Hotel names were associated with OpenStreetMap data by using Photon in a Samarkand boundary. First, candidate names were produced and then only reasonable matches were selected. Coordinates which were exact duplicates on the same platform were considered uncertain; the name with the highest score was chosen. Moreover, the four remote matches more than 12 km away from Registan and those inconsistent with the Registan-distance field in Ctrip by more than 1.2 km were excluded. As a result, a spatial sample of 171 hotels was obtained, which is fewer than the attribute sample but less susceptible to false precision.

Table 1. Ctrip data structure and analytical samples

Layer	Records	Key fields	Quality treatment
Hotel attributes	323	Rating, review count, four sub-scores, POI distances	21 missing ratings retained only for count-based analysis
Guest reviews	4,661	Score, text, travel party, stay month, photos	No exact row duplication detected

Layer	Records	Key fields	Quality treatment
Initial geocodes	278	Longitude and latitude candidates	Samarkand boundary and name-score checks
Conservative spatial sample	171	Unique validated coordinates	One-to-one coordinate retention and outlier filtering

Source: author calculations from the platform files supplied for this study.

The point pattern is summarized by the average nearest-neighbour index (NNI), a Gaussian kernel-density surface, the mean center, a two-standard-deviation ellipse, and distance rings from 0 to 1, 1 to 2, 2 to 3, 3 to 5 and more than 5 km away from Registan. The NNI indicates the difference between the actual mean nearest-neighbour distance and the expected value in complete spatial randomness within the convex hull of the verified points.

For hotel ratings and model residuals, global Moran's I is calculated with a row-standardized eight-nearest-neighbour matrix and 999 random permutations. The econometric result is ln(1 + hotel review count). A maximum-likelihood spatial autoregressive model involves standardized hotel rating, geodesic distance to Registan, and Ctrip-reported driving distances to the airport and railway station.

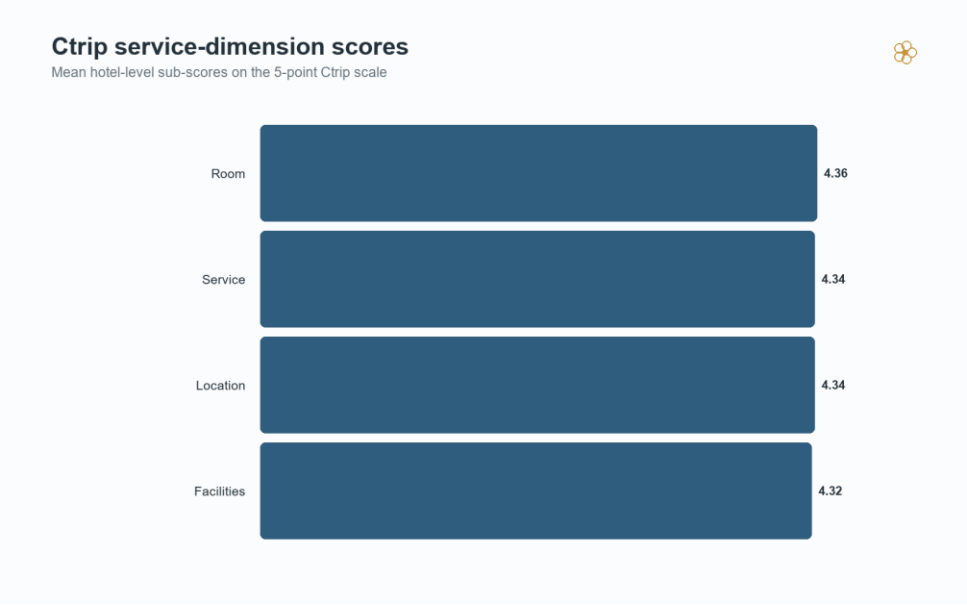
The spatial coefficient ρ represents the conditional relationship between the attention of a hotel's website and that of its neighboring hotels. Since the data are cross-sectional, it is considered as a spillover effect rather than indicating that one hotel affects the reviews of another. The theme incidence is determined by the ratio of review texts containing at least one multi-lingual term in a given service dictionary to the total number of reviews, not to the actual number of tokens.

Results

The average Ctrip score of the hotels is 4.34 (SD 0.61) on the five-point scale. The sub-scores are closely concentrated: room 4.36, service 4.34, location 4.34 and facilities 4.32. The small variation suggests that minor numerical changes should not be considered as important practically.

Figure 1. Mean Ctrip hotel sub-scores

Source: Ctrip hotel sheet; n=302 rated properties.



The verified point distribution is not random. The observed minimum distance to its nearest neighbour is 0.237 km, whereas it should be 0.323 km if it were random, resulting in NNI=0.734 (z=-6.66, p<0.001). The kernel density reaches its maximum only 0.12 km away from Registan. The mean centre is located at 66.9668° E, 39.6568° N; the elongated ellipse and the apparent eastern satellite group indicate the coexistence of the old-town facilities and the Silk Road resort centre.

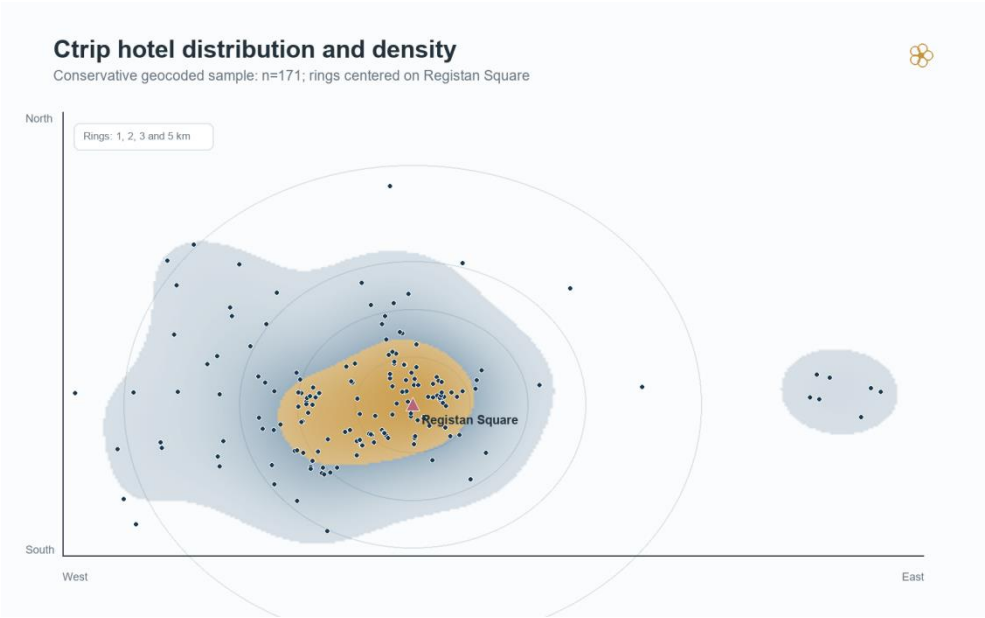


Figure 2. Ctrip hotel distribution and kernel density

Source: author geocoding and spatial calculations; conservative one-to-one sample n=171.

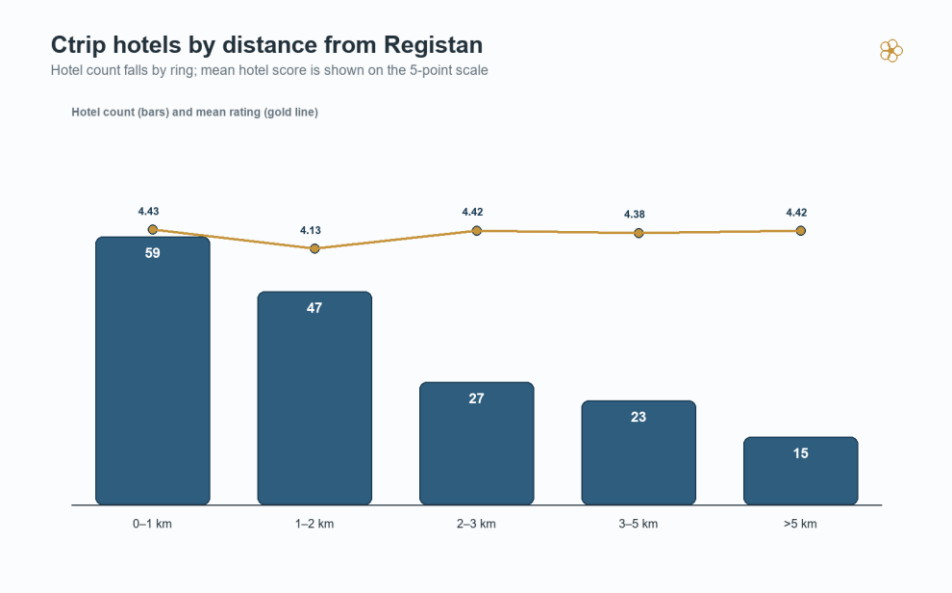
Table 2. Ctrip spatial-statistical results

Statistic	Estimate	Inference
Nearest-neighbour index	0.734	z=-6.66; p<0.001
Convex-hull area	71.43 km²	Validated point envelope
KDE peak distance to Registan	0.12 km	Bandwidth=0.79 km
Rating Moran's I	-0.005	Permutation p=0.882
Rating–distance Spearman rho	-0.020	Approx. p=0.803

Source: author calculations from the platform files supplied for this study.

Among some specific places, 59 (34.5%) are within 1 km of Registan and 106 (62.0%) are within 2 km. The numbers decrease in the next rings, but the ratings do not show a monotonic decrease: the means of the rings range from 4.13 to 4.43. The almost no correlation between rating and distance, as well as the Moran's statistic, suggest that the geographical distribution of the quality of Ctrip is generally neutral, although the supply distribution is different.

Figure 3. Ctrip hotel counts and ratings by Registan distance ring



Source: conservative geocoded Ctrip sample; rating means use non-missing scores.

Table 3. Ctrip ring-zone profile

Ring	Hotels	Share	Meanrating	Platform reviews
0–1 km	59	34.5%	4.43	1,299
1–2 km	47	27.5%	4.13	926
2–3 km	27	15.8%	4.42	562
3–5 km	23	13.5%	4.38	236
>5 km	15	8.8%	4.42	967

Source: author calculations from the platform files supplied for this study.

The SAR model is based on 147 hotels with all model fields. The spatial coefficient is positive and statistically significant (rho=0.43; likelihood-ratio p=0.003). This is the most clear result specific to the platform: the accumulation of reviews is locally correlated after controlling for rating and accessibility. The standardized rating coefficient is 0.282 (p=0.021), while the direct Registan-distance coefficient is not significant at 5%. The residual Moran statistic is -0.016 (permutation p=0.702), suggesting that the fitted spatial process takes up the remaining local dependence.

Table 4. Spatial-lag model of Ctrip review volume

Term	Coefficient	SE	zz	pp
Hotel rating (z)	0.282	0.122	2.32	0.021
Distance to Registan (z)	0.197	0.145	1.36	0.175
Airport drive distance (z)	-0.123	0.150	-0.82	0.411
Rail-station drive distance (z)	0.122	0.142	0.86	0.390
Spatial lag (rho)	0.430	—	LR test	0.003

Source: author calculations from the platform files supplied for this study.

The 4,661 comments have an average score of 4.54 (SD 0.83) with a median of 5. The recognition of Latin-script is found in 34.2% of the comments, Chinese script in 30.4%, Cyrillic in 24.9%, and explicit Japanese kana in 8.5%. This suggests that the Ctrip corpus includes readers from various East-Asian countries rather than a single Chinese-speaking group.

At the review level, room and cleanliness terms occur in 34.6% of texts, followed by service and staff (29.6%) and breakfast and dining (27.7%). Location and transport account for 19.9%. This percentage-based result is contrary to the preliminary ranking by raw keyword frequency: repeated location words may have a high token count but do not necessarily appear in the majority of individual reviews. Therefore, the new method is more appropriate for evaluating the variety of topics.

Figure 4. Incidence of guest-experience themes in Ctrip reviews
Source: multilingual dictionary coding of 4,661 review texts; topics are non-exclusive.

Travelers in different groups are found; alone travelers account for 19.9%, couples 18.6%, families 17.4%, friends 15.5% and business men/women 11.2%. The number of reviews is mostly concentrated in April-May and September-October, but these data should be regarded as the monthly distribution of the stayed places rather than the occupancy rate, since the review periods involve various years.

Discussion

The Ctrip market includes three spatial processes. The hotel supply is fixed at certain monuments; the hotel quality is unaffected by the geographical location; and the platform popularity has a local influence. The positive rho is in accordance with the neighborhood-level demand cycles and search visibility. When a tourist selects an old-town hotel with good reviews, he is also exposed to other nearby alternatives, and the local accessibility among different hotels is the same. In the resort area, branded hotels also gain advantage from a unified destination product. Therefore, the spatial popularity cannot be simply explained by the distance from a single monument.

The quality neutrality result is also important. Being near Registan affects the location of the hotels, but not the ones with high scores. The facilities and services are the main reasons for the competition in various aspects. The topic of the reviews confirm this conclusion: the number of reviews about the room condition, staff interaction and breakfast is larger than that about the location when the occurrence is taken into account instead of the total number of words.

For operators facing Ctrip, the use of multilingual content is necessary. It is required to have property descriptions, arrival instructions and service recoveries in several languages. The Japanese-script part should be paid special attention, as it is displayed in the Chinese platform channel.

Managerial and planning implications

Destination managers should consider dense hotel neighborhoods as a common platform system. The improvement of walking routes, wayfinding, pick-up zones and shuttle services can increase the attractiveness of several adjacent hotels at the same time. Furthermore, the extensive spatial spill-over indicates that local reputation problems may also influence the surrounding substitutes; hence, it is necessary to establish area-wide quality standards and coordinate visitor information.

Hotels near Registan should not depend solely on its location. The lack of a rating difference and the assessment of the room, service and breakfast indicate that the benefit of being in an historical area is doubtful. The hotels situated in the suburbs or as resorts can make up for the distance by

providing satisfactory service and convenient transportation, whereas the central hotels must enhance the cleanliness and multilingual reception to maintain their good reputation.

Limitations

The analysis is based on one photograph of a certain website and a non-systematically geocoded sample. The hotels with names resembling OpenStreetMap might be larger or more widely known compared to those which do not match. The total number of reviews is complete but does not distinguish the accommodation capacity, the time period on the platform and the present demand. The SAR model is only correlational; it can not determine the influence of peer effects. The multilingual dictionary merely counts the occurrences of topics without considering the sentiment polarity, irony or intensity of topics. Additional research should take several photographs, obtain the actual room capacity and prices, and estimate the hierarchical spatial models with house fixed effects.

Conclusion

Ctrip's Samarkand market is located around Registan, but its quality indicator is spatially unbiased. The main discovery is a positive spatial correlation in the number of reviews (rho=0.43), indicating that the platform's attention concentrates at the local level. When considering the multilingual review system and the predominance of basic service themes, the conclusion changes from 'how near the hotel is to the landmark?' to 'how well the hotel transforms a common location into a reliable and multilingual guest experience and review influence?'.

REFERENCES

1. Anselin, L. (1988). Spatial Econometrics: Methods and Models. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-015-7799-1
2. Anselin, L. (1995). Local indicators of spatial association—LISA. Geographical Analysis, 27(2), 93–115.
3. Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. Geographical Analysis, 24(3), 189–206.
4. Hou, Z., Cui, F., Meng, Y., Lian, T., & Yu, C. (2019). Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis. Tourism Management, 74, 276–289. https://doi.org/10.1016/j.tourman.2019.03.009
5. Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. Tourism Management, 29(3), 458–468.
6. Ma, Y., Li, H., & Tong, Y. (2022). Distribution differentiation and influencing factors of the high-quality development of the hotel industry. Sustainability, 14(11), 6476.
7. Moran, P. A. P. (1948). The interpretation of statistical maps. Journal of the Royal Statistical Society: Series B, 10(2), 243–251.
8. OpenStreetMap contributors. (2026). OpenStreetMap data for Samarkand. https://www.openstreetmap.org
9. Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. Tourism Management, 58, 51–65. https://doi.org/10.1016/j.tourman.2016.10.001
10. Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? International Journal of Hospitality Management, 44, 120–130. https://doi.org/10.1016/j.ijhm.2014.10.013