

Online Platform Data for Tourism Research in Northern Albania: A Comparative Analysis of Booking.com and TripAdvisor

Fatjona Kroni

University of Shkodra, Faculty of Natural Sciences
fatjona.kroni@unishk.edu.al

Ermira Kalaj

University of Shkodra, Faculty of Economics
ermira.kalaj@unishk.edu.al

Erjola Barbullushi

University of Shkodra, Faculty of Economics
erjola.barbullushi@unishk.edu.al

Abstract: The increasing digitalization of the tourism sector has expanded the availability of data from online travel platforms, creating new opportunities for tourism research, particularly in destinations where traditional statistics are limited. Building on existing literature that highlights the growing role of user-generated content and platform-based data, this study examines the potential of online platforms as sources of tourism insights. A comparative analysis is conducted using two datasets collected from Booking.com and TripAdvisor, focusing on accommodation listings in Northern Albania. Data preprocessing and analysis were performed using Python, enabling the cleaning, matching, and normalization of accommodation units across platforms. The study applies descriptive statistics and cross-platform comparison to explore differences in rating distribution, review volume, and how accommodation performance is represented. The findings indicate that while both platforms reflect similar overall trends, significant differences emerge in the variability and interpretation of key indicators. These differences are linked to platform-specific characteristics, including user participation and evaluation mechanisms. The study highlights the importance of interpreting online platform data and demonstrates the value of comparative approaches for tourism research in emerging destinations.

Keywords: Online platform data; Booking.com; TripAdvisor; Comparative analysis; Northern Albania

1 Introduction

The increasing digitalization of the tourism sector has transformed the way travelers search for, evaluate, and select accommodation services. Platforms such as Booking.com and TripAdvisor have become central sources of information, providing access to user-generated content, including ratings, reviews, and descriptive details about accommodation units. Beyond their role as information channels, these platforms have also emerged as key sources of digital data, generating large volumes of information that can complement traditional tourism statistics and support data-driven analysis [1].

In addition to their informational role, online platforms have gained importance as valuable data sources for tourism research. The availability of platform-based data offers new opportunities to analyze tourism performance, particularly in destinations where traditional statistical data may be limited, aggregated, or delayed [2]. In the context of Northern Albania, official tourism statistics often do not fully capture patterns of accommodation performance, user engagement, and digital visibility. As a result, data derived from online platforms can provide complementary insights into the dynamics of the tourism sector [3].

However, online platform data are not neutral. Differences in platform structure, user behavior, review verification systems, and evaluation mechanisms can influence how information is generated and represented. Booking.com, for example, collects reviews only from verified users who have completed a stay, while TripAdvisor allows broader participation. These differences may affect rating consistency, review volume, and the representation of accommodation performance [4]. Consequently, data obtained from different platforms may not be directly comparable.

Existing research highlights the importance of analyzing multiple platforms to better understand these differences. While many studies have relied on a single data source, comparative approaches allow researchers to identify inconsistencies, biases, and platform-specific patterns in online tourism data [5]. This is particularly important when platform-based indicators are used to support tourism analysis, decision-making, or policy development.

In this context, the present study aims to examine the potential of online platform data for tourism research through a comparative analysis of Booking.com and TripAdvisor accommodation listings in Northern Albania. The study focuses on accommodation units present on both platforms and compares key indicators, including rating scores, review volume, and rating distribution patterns. Using

Python for data preprocessing, matching, and analysis, the study applies descriptive statistics and visual exploration to identify similarities and differences in how tourism performance is represented across platforms [6].

The findings of this study provide insights into the extent to which online platforms reflect consistent or divergent representations of accommodation performance. By highlighting differences in user engagement, rating variability, and platform-based evaluation systems, the study contributes to a better understanding of the strengths and limitations of online platform data. The research emphasizes the importance of interpreting such data when used for tourism analysis, particularly in emerging destinations such as Northern Albania.

2 Literature review

The development of Information and Communication Technologies (ICTs) has fundamentally transformed the tourism industry, giving rise to the concept of eTourism. ICTs have reshaped business models, enabling more efficient distribution systems, improving service delivery, and facilitating direct interaction between tourism stakeholders and consumers [7]. This transformation has accelerated with the emergence of digital platforms, which now act as central intermediaries in the tourism ecosystem.

Digital platforms such as Booking.com and TripAdvisor function as multi-sided marketplaces that connect service providers and consumers while reducing search costs and increasing market transparency [8]. These platforms enable tourists to access, evaluate, and compare accommodation options easily, thus influencing travel decision-making processes. The integration of user-generated content (UGC) has further strengthened their role, making them key sources of tourism-related information.

Online reviews, as a form of electronic word-of-mouth (eWOM), play a crucial role in reducing uncertainty associated with tourism products, which are inherently intangible and experience-based. Travelers rely on peer-generated content to evaluate service quality and make informed decisions. Research shows that online reviews significantly influence booking behavior and contribute to building trust and reducing perceived risk [9]. Moreover, online reviews not only influence consumers but also provide valuable feedback to service providers, helping them improve performance and competitiveness [10].

Despite their importance, online review platforms are not homogeneous. Significant differences exist between platforms such as TripAdvisor and Booking.com in terms of data structure, user behavior, and evaluation mechanisms. Booking.com relies on verified reviews from actual customers, whereas TripAdvisor allows broader participation, including anonymous contributions. These structural differences

influence rating patterns, review content, and data reliability, leading to variations in how accommodation performance is represented [11].

Beyond structural differences, platforms also vary in how they shape user perceptions and represent accommodation experiences. While Booking.com provides a more standardized evaluation framework, TripAdvisor tends to reflect a wider range of subjective experiences. As a result, the same accommodation unit may be perceived differently depending on the platform, highlighting the importance of examining multiple sources when analyzing tourism data [12].

Research comparing these platforms highlights important differences. For instance, the relationship between the number of reviews and rating scores differs significantly between platforms. A positive relationship between review volume and scores is often observed on TripAdvisor, while this relationship is weaker or absent on Booking.com [13]. This suggests that platform dynamics, such as user participation and rating systems, directly affect how aggregated indicators are formed.

The increasing availability of digital data has further expanded opportunities for tourism research. Online platforms generate vast amounts of real-time data, offering more granular and timely insights compared to traditional statistics. Big data sources, including social media and online booking platforms, provide new opportunities for measuring tourism flows, performance, and consumer behavior, although they also introduce challenges related to data quality and bias [1].

In this context, web scraping has become an important methodological tool for extracting and analyzing large-scale tourism data. It allows researchers to collect structured and unstructured data from online platforms efficiently, enabling deeper insights into customer behavior, preferences, and market dynamics [14]. Such approaches are particularly valuable in regions where official tourism statistics are limited or delayed.

In the case of Albania, the use of online platform data presents a significant opportunity for tourism analysis. Studies have demonstrated that data from platforms such as TripAdvisor can be used to develop indicators of tourism development, offering insights into accommodation quality, customer satisfaction, and market trends [3]. This is particularly relevant for emerging destinations, where traditional data sources may not fully capture tourism dynamics.

The literature demonstrates that online travel platforms are not neutral data sources but are shaped by platform-specific characteristics, including verification systems, user behavior, and evaluation mechanisms. These differences influence how tourism performance is represented and interpreted. A comparative approach is necessary to better understand the extent to which platform-based data provide consistent and reliable insights, particularly in emerging destinations such as Northern Albania.

3 Methodology

This study applies an exploratory, data-driven methodology to compare how two major online travel platforms, Booking.com and TripAdvisor, represent accommodation performance in Northern Albania. The methodological design follows prior comparative platform studies, which emphasize the need to analyze the same accommodation units across different online review systems in order to identify structural differences in rating behavior, review volume, and platform-based representation.

The study is based on two datasets collected from Booking.com and TripAdvisor. The datasets include accommodation listings located in Northern Albania and contain key variables such as accommodation name, rating score, number of reviews, and descriptive attributes related to each accommodation unit. Online platform data are used as an alternative and complementary source of tourism information, particularly in contexts where official statistics may be limited in granularity, frequency, or detail [15].

Data preprocessing and analysis were conducted using Python. The preprocessing phase included cleaning the datasets, standardizing accommodation names, checking for missing or inconsistent values, and preparing the variables for comparison. This follows the logic of exploratory data analysis, where Python is used to inspect datasets, identify missing values, clean incorrect records, examine variable distributions, and visualize relationships between variables [6][14].

To identify accommodation units present on both platforms we applied a matching procedure. Since hotel names may appear with small variations across platforms, the matching process involved standardizing names, removing unnecessary characters, checking spelling differences, and manually verifying uncertain matches. Only matched accommodation units were retained for the final comparative analysis. This step was necessary because the objective of the study is not to compare the platforms in general, but to examine how the same accommodation units are represented differently across Booking.com and TripAdvisor.

To ensure comparability, TripAdvisor ratings were transformed to a common 0–10 scale, aligning them with the rating scale used by Booking.com. This normalization step is important because previous studies show that differences in platform rating scales can influence the interpretation of hotel performance across Booking.com and TripAdvisor [11]. Review counts were retained as reported on each platform, as they represent the level of user engagement and digital visibility generated by each accommodation unit.

The analytical approach combines descriptive statistics, visual exploration, and cross-platform comparison. First, review counts were aggregated by platform to compare the level of user engagement. Second, rating distributions were examined

to assess differences in variability and concentration across platforms. Third, the relationship between Booking.com and TripAdvisor ratings was analyzed using a scatter plot with a fitted trend line, allowing the study to identify whether accommodations rated highly on one platform also tend to receive high ratings on the other. This approach is consistent with comparative platform analytics, where multiple online review platforms are examined to understand whether they provide consistent or divergent representations of hotel performance.

This study is exploratory in nature rather than predictive. It aims to identify patterns, differences, and platform-specific tendencies, without seeking to establish generalizable causal relationships. This is appropriate because online platform data may be affected by platform design, user behavior, data availability, and self-selection bias. The findings are interpreted cautiously as evidence of cross-platform differences in the representation of accommodation performance in Northern Albania.

4 Results

4.1 Review volume across platforms

The comparative analysis of review counts reveals a clear difference in user engagement between the two platforms. Booking.com records 6,449 reviews compared to 1,092 on TripAdvisor, indicating a substantially higher level of user participation. This difference suggests that review volume varies considerably across platforms and should be interpreted with caution in comparative analyses.

4.2 Rating distribution across platforms

The distribution of ratings across platforms reveals notable differences in variability and consistency. As shown in Figure 1, ratings on Booking.com are relatively concentrated within a narrow range, indicating a more stable and consistent evaluation pattern among users. The red line represents the mean rating for each platform, highlighting the difference in central tendency between Booking.com and TripAdvisor. Most ratings are clustered at higher values, suggesting limited dispersion.

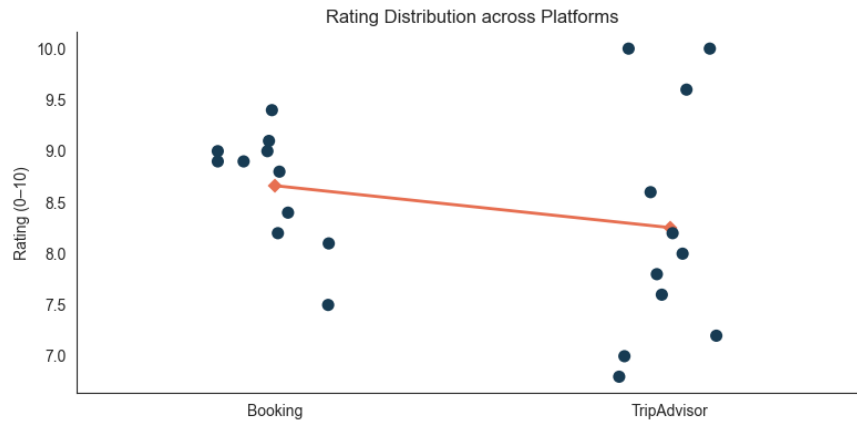


Figure 1
Distribution of accommodation ratings across Booking.com and TripAdvisor

In contrast, TripAdvisor ratings display a wider spread. This variability indicates more diverse user evaluations. The observed differences suggest that the two platforms capture user sentiment in distinct ways, likely influenced by differences in review systems and user behaviour.

4.3 Relationship between ratings across platforms

The relationship between ratings on Booking.com and TripAdvisor is presented in Figure 2 through a scatter plot. The results indicate a positive association between ratings across the two platforms, suggesting that accommodations with higher ratings on one platform tend to receive higher ratings on the other. The red line represents the regression trend, showing the overall relationship between the two variables. However, the dispersion of data points around the trend line reveals that this relationship is not perfectly aligned. Several accommodations exhibit noticeable differences in ratings between platforms, indicating that the same unit may be evaluated differently depending on the platform. This variability highlights the presence of platform-specific factors influencing user evaluations.

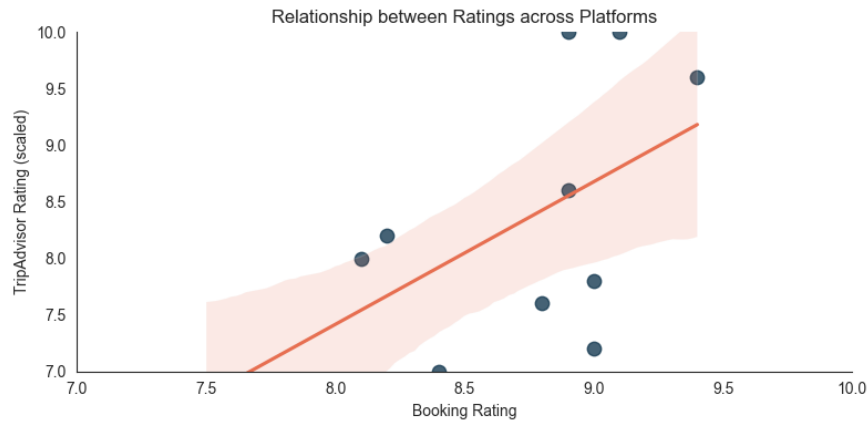


Figure 2
Relationship between accommodation ratings on Booking.com and TripAdvisor

While a general trend of consistency is observed, the differences at the individual level suggest that platform-based ratings should not be considered directly interchangeable.

5 Discussion

The findings of this study confirm that online travel platforms differ not only in structure but also in how they represent tourism performance. The comparative analysis of Booking.com and TripAdvisor data reveals both similarities and differences in key indicators, highlighting the need for careful interpretation of platform-based information.

A key result of the analysis is the difference in review volume between the two platforms, with Booking.com showing much higher user engagement. This aligns with prior research suggesting that review volume is associated with the stability and representation of ratings across online platforms [13]. A larger number of reviews increases the influence of online information and reduces uncertainty in consumer decision-making [16]. However, the imbalance in review counts suggests that the two platforms operate under different engagement dynamics, which may affect data representativeness.

The analysis of rating distribution reveals that Booking.com ratings are more concentrated, while TripAdvisor ratings display greater variability. This reflects structural differences between platforms, particularly in their review verification mechanisms. Booking.com restricts reviews to verified users who have completed

a stay, leading to more stable and consistent rating patterns. In contrast, TripAdvisor allows broader participation, resulting in more diverse and potentially more volatile evaluations [4]. The wider dispersion of TripAdvisor ratings observed in this study supports this interpretation and confirms that platform design plays a crucial role in shaping user evaluations. This is consistent with previous comparative studies on Booking.com and TripAdvisor, which show that differences in platform structure and user participation can lead to divergent representations of accommodation performance [17].

The relationship between ratings across platforms is positive but not identical. While accommodations with higher ratings on one platform tend to receive higher ratings on the other, deviations are observed. This indicates that ratings are influenced by platform-specific factors and should not be treated as directly interchangeable. This finding is in line with research suggesting that rating behavior is not always symmetric and may vary depending on context and platform characteristics [18].

These differences can be explained by how platforms influence consumer perception. Although both Booking.com and TripAdvisor rely on user-generated content, they present accommodation experiences in different ways. TripAdvisor reflects a broader range of subjective experiences, while Booking.com provides a more structured and standardized evaluation framework. The results of this study support this perspective, as the greater variability in TripAdvisor ratings suggests a wider range of user opinions and experiences [12].

The results show that online travel platforms should not be treated as directly equivalent data sources. Differences in user engagement, rating patterns, and platform mechanisms affect how tourism performance is represented. These findings are particularly relevant for emerging destinations such as Northern Albania, where online platforms can serve as key sources of tourism insights [3]. Researchers and practitioners should therefore consider platform-specific characteristics when analyzing and comparing tourism performance indicators.

Conclusions

This study examined the use of online travel platforms as sources of tourism data through a comparative analysis of Booking.com and TripAdvisor in Northern Albania. The results show that, while both platforms reflect similar general trends, they differ in review volume, rating distribution, and the way accommodation performance is represented. These differences are related to platform-specific features, including review systems and user participation.

The findings indicate that data from online platforms should not be treated as directly comparable and require careful interpretation when used for analysis or decision-making. The study also shows that a comparative approach helps to

identify platform-specific differences and provides a clearer understanding of how tourism data are represented.

For emerging destinations such as Northern Albania, online platforms can serve as a useful complement to traditional statistics, although their limitations must be considered. Future research may extend this work by including additional platforms, larger datasets, or more advanced analytical methods to further examine the dynamics of platform-based tourism data.

References

- [1] Christophe Demunter: Tourism Statistics: Early Adopters of Big Data?, Publications Office of the European Union, Luxembourg, 2017. Available at: <https://ec.europa.eu/eurostat/documents/3888793/8234206/KS-TC-17-004-EN-N.pdf>
- [2] Marta Drozdowska, Magdalena Duda-Seifert: Travel Websites: A Relevant Source of Statistical Information?, *Turyzm/Tourism*, Vol. 26, No. 2, 2016, pp. 7–13, DOI: 10.1515/tour-2016-0008.
- [3] Brunilda Liçaj, Leida Matja: Use of Data Web on Trip Advisor as an Opportunity to Develop Indicators of Tourism Development in Albania, in *Proceedings of the 13th International Conference of ASECU: Social and Economic Challenges in Europe 2016–2020*, Durrës, Albania, May 19–20, 2017.
- [4] Paolo Figini, Laura Vici, Giampaolo Viglia: A Comparison of Hotel Ratings Between Verified and Non-Verified Online Review Platforms, *International Journal of Culture, Tourism and Hospitality Research*, Vol. 14, No. 2, 2020, pp. 157–171, DOI: 10.1108/IJCTHR-10-2019-0193.
- [5] Zheng Xiang, Qianzhou Du, Yufeng Ma, Weiguo Fan: A Comparative Analysis of Major Online Review Platforms: Implications for Social Media Analytics in Hospitality and Tourism, *Tourism Management*, Vol. 58, 2017, pp. 51–65, DOI: 10.1016/j.tourman.2016.10.001.
- [6] Kabita Sahoo, Abhaya Kumar Samal, Jitendra Pramanik, Subhendu Kumar Pani: Exploratory Data Analysis using Python, *International Journal of Innovative Technology and Exploring Engineering*, Vol. 8, No. 12, 2019, pp. 4727–4735, DOI: 10.35940/IJITEE.L3591.1081219.
- [7] Dimitrios Buhalis, Rob Law: Progress in Information Technology and Tourism Management: 20 Years on and 10 Years after the Internet—The State of eTourism Research, *Tourism Management*, Vol. 29, No. 4, 2008, pp. 609–623, DOI: 10.1016/j.tourman.2008.01.005.
- [8] Tatiana Genzorová, Tatiana Čorejová, Natália Stalmašeková: Comparing the Use of Digital Platforms in Tourism, in *Proceedings of CBU International*

- Conference 2018: Innovations in Science and Education, Prague, Czech Republic, 2018, Vol. 6, pp. 152–155, DOI: 10.12955/cbup.v6.1148.
- [9] Zhiwei Liu, Sangwon Park: What Makes a Useful Online Review? Implication for Travel Product Websites, *Tourism Management*, Vol. 47, 2015, pp. 140–151, DOI: 10.1016/j.tourman.2014.09.020.
 - [10] Jie J. Zhang, Rohit Verma: What Matters Most to Your Guests: An Exploratory Study of Online Reviews, *Cornell Hospitality Report*, Vol. 17, No. 4, Cornell University, Ithaca, New York, USA, 2017. Available at: <https://sha.cornell.edu/wp-content/uploads/sites/4/2019/03/rohit-verma-online-reviews.pdf>
 - [11] Eva Martin-Fuentes, Carles Mateu, Cesar Fernandez: Does Verifying Uses Influence Rankings? Analyzing Booking.Com and Tripadvisor, *Tourism Analysis*, Vol. 23, No. 1, 2018, pp. 1–15, DOI: 10.3727/108354218X15143857349459.
 - [12] Maria Teresa Borges-Tiago, Carolina Arruda, Flávio Gomes Borges Tiago, Paula Rita: Differences Between TripAdvisor and Booking.com in Branding Co-Creation, *Journal of Business Research*, Vol. 123, 2021, pp. 380–388, DOI: 10.1016/j.jbusres.2020.09.050.
 - [13] Eva Martin-Fuentes, Carles Mateu, Cesar Fernandez: The More the Merrier? Number of Reviews Versus Score on TripAdvisor and Booking.com, *International Journal of Hospitality and Tourism Administration*, Vol. 21, No. 1, 2020, pp. 1–14, DOI: 10.1080/15256480.2018.1429337.
 - [14] Gianluca Barbera, Luis Araujo, Silvia Fernandes: The Value of Web Data Scraping: An Application to TripAdvisor, Big Data and Cognitive Computing, Vol. 7, No. 3, 2023, Article 121, DOI: 10.3390/bdcc7030121.
 - [15] Cristina Rodica Boboc, Ana Maria Babaligea, Simona Ioana Ghiță, Andreea Simona Săseanu: Leveraging Web-Scraping for Tourism Data Analysis: A Case Study on Romania, in *New Trends in Sustainable Business and Consumption*, BASIQ International Conference, Bucharest, Romania, 2025, pp. 174–181, DOI: 10.24818/BASIQ/2025/11/020.
 - [16] Wenjing Duan, Bin Gu, Andrew B. Whinston: Do Online Reviews Matter? —An Empirical Investigation of Panel Data, *Decision Support Systems*, Vol. 45, No. 4, 2008, pp. 1007–1016, DOI: 10.1016/j.dss.2008.04.001.
 - [17] Oktay Emir, Ramazan Ufuk Bayer: Examination of Urban Hotels' Online Review Site Performance with a Comparative Approach Based on TripAdvisor and Booking.com, *Journal of Business Research-Türk*, Vol. 9, No. 3, 2017, pp. 72–85, DOI: 10.20491/isarder.2017.288.
 - [18] Lawrence Hoc Nang Fong, Soey Sut Ieng Lei, Rob Law: Asymmetry of Hotel Ratings on TripAdvisor: Evidence from Single- Versus Dual-Valence Reviews, *Journal of Hospitality Marketing & Management*, Vol. 26, No. 1, 2017, pp. 67–82, DOI: 10.1080/19368623.2016.1178619.