

Title: Leveraging Web-scraping for Tourism Data Analysis: a Case Study on Romania

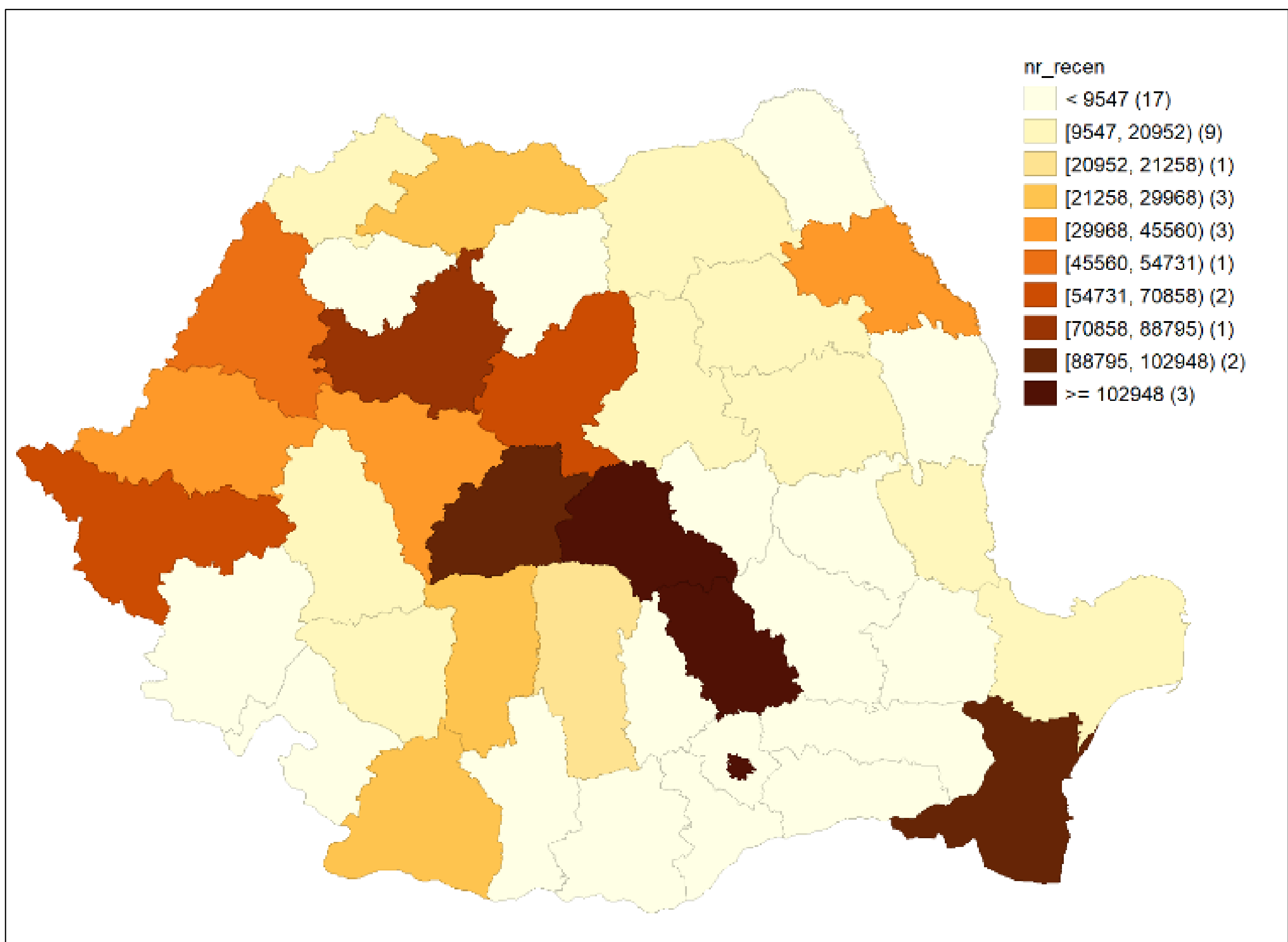
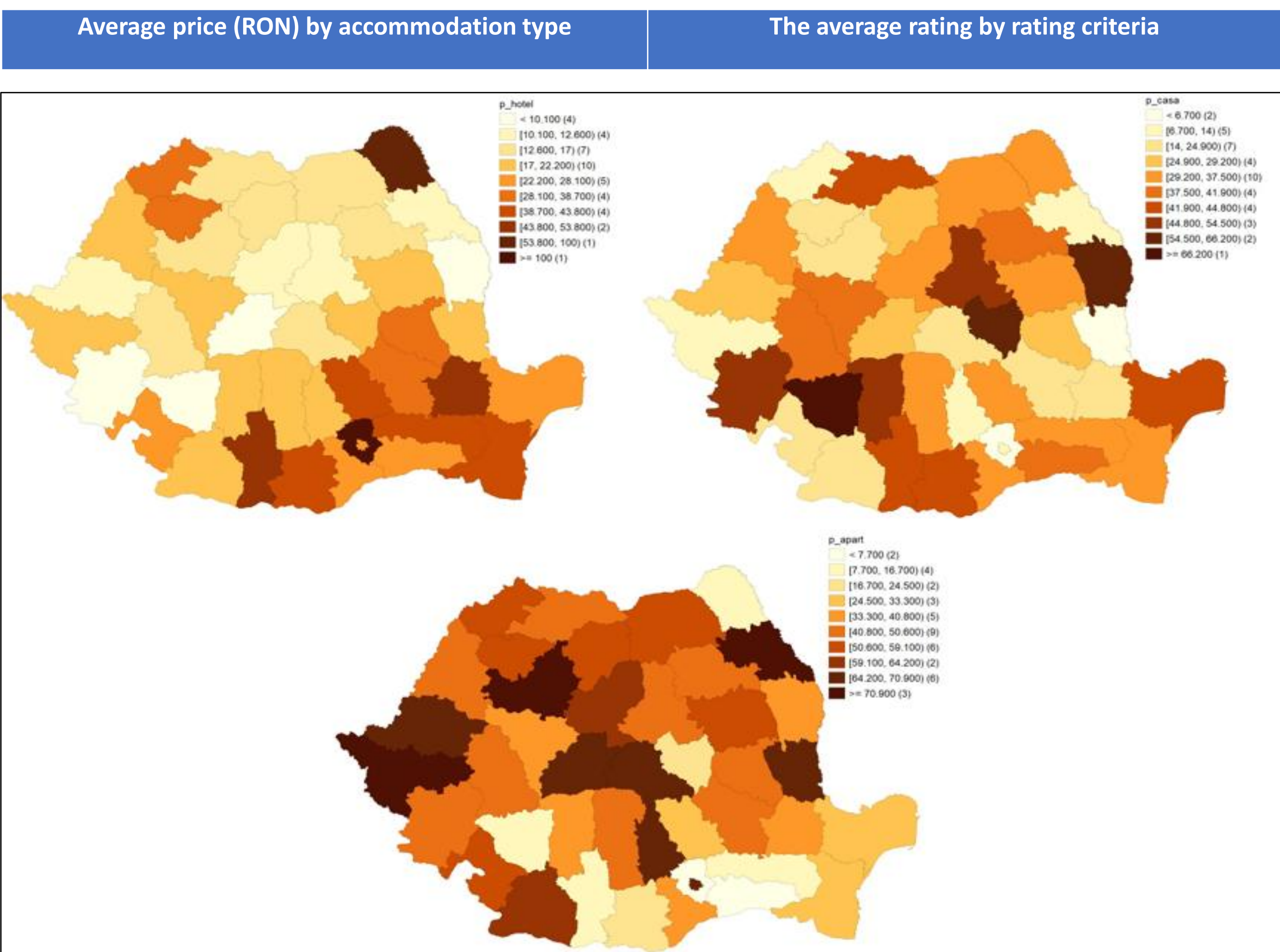
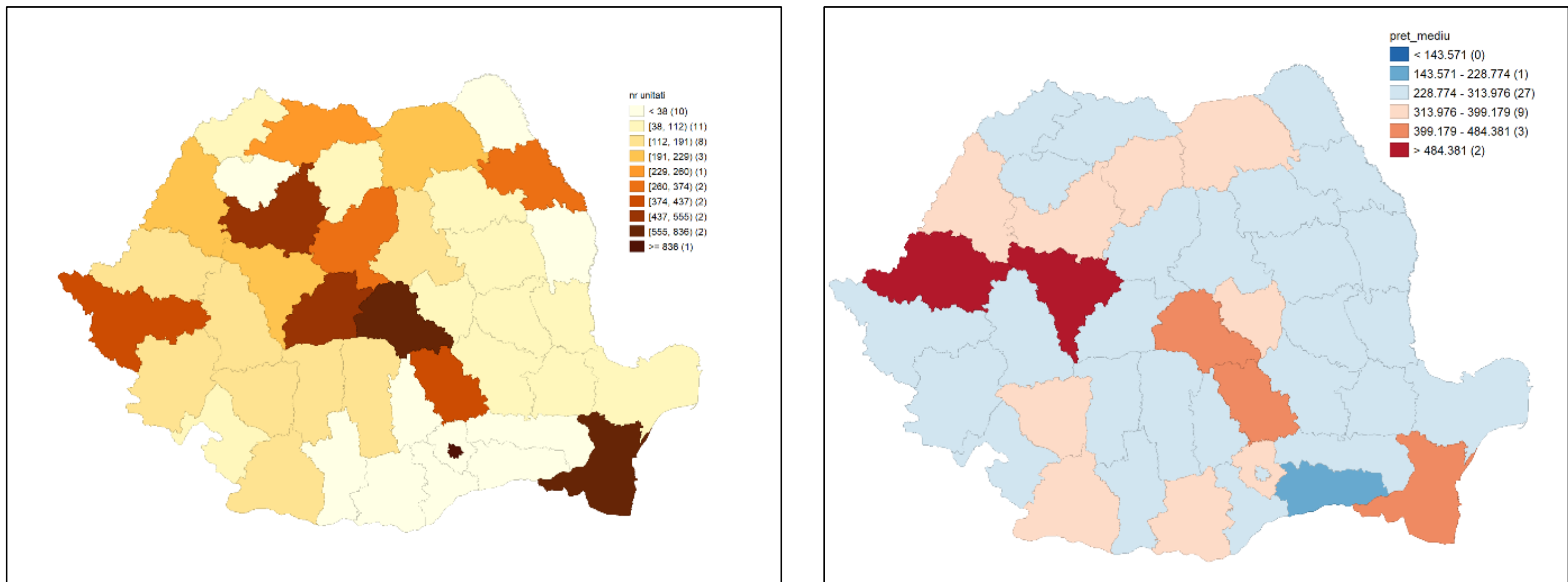
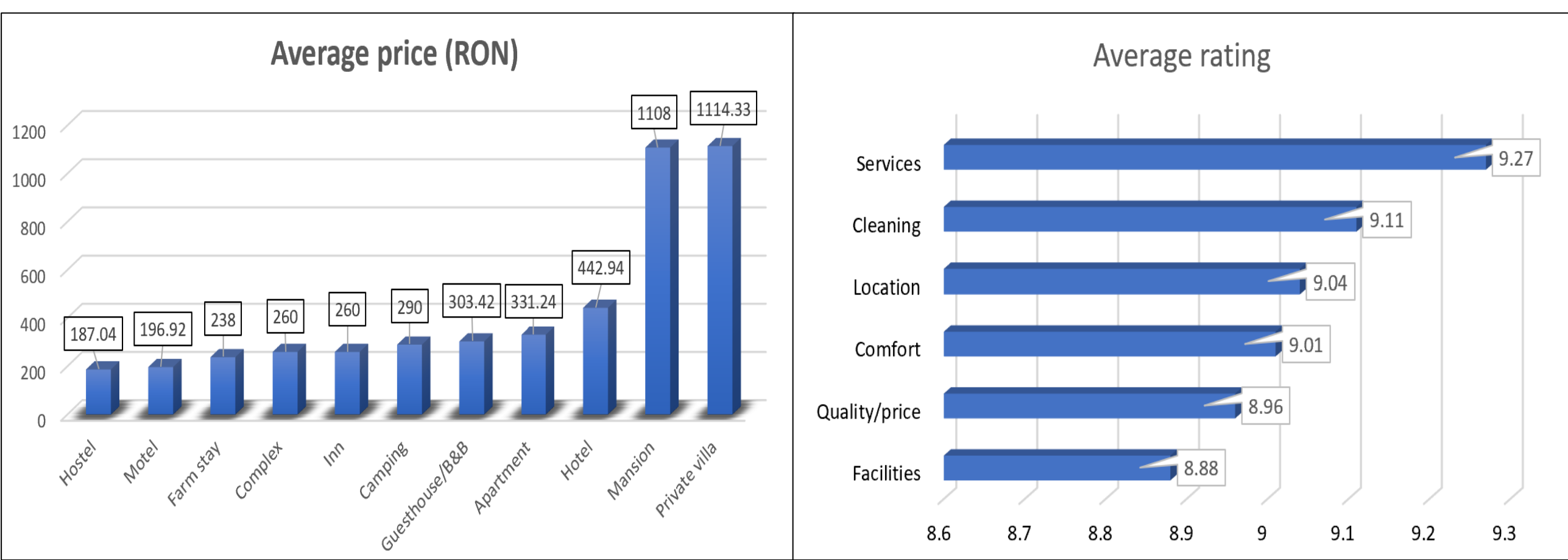
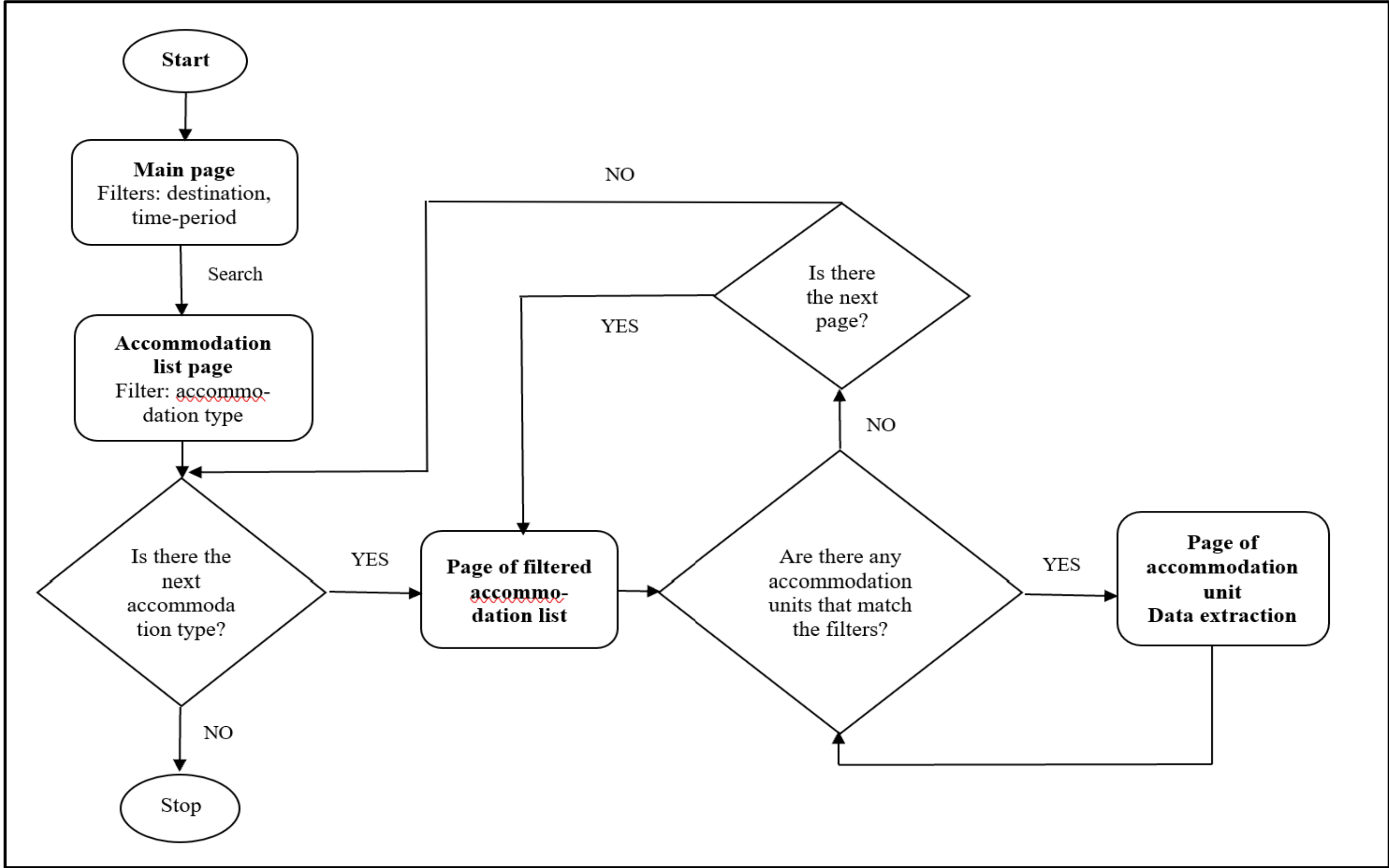
Authors: Cristina Rodica Boboc<sup>1</sup>, Ana Maria Babaligea<sup>2</sup>, Simona Ioana Ghiță<sup>3</sup> and Andreea Simona Săseanu<sup>4</sup>

Affiliation: <sup>1)2)3)4)</sup> Bucharest University of Economic Studies, Bucharest, Romania. <sup>1)3)</sup> Institute of National Economy, Bucharest, Romania

Email: [cristina.boboc@csie.ase.ro](mailto:cristina.boboc@csie.ase.ro); [babaligeaana19@stud.ase.ro](mailto:babaligeaana19@stud.ase.ro); [simona.ghita@csie.ase.ro](mailto:simona.ghita@csie.ase.ro); [andreea.saseanu@com.ase.ro](mailto:andreea.saseanu@com.ase.ro)

Abstract

In the digitalization era, the availability and accessibility of data have experienced a significant increase, opening up new opportunities for the analysis of the tourism sector—an area where the primary source of data for performance evaluation has traditionally been official statistics. This paper aims to investigate the potential of using an alternative data source, namely web scraping, by emphasizing the additional insights and advantages that this method can offer in comparison to conventional statistical data. The information was collected through web scraping from an online tourism booking platform, using a specially developed program written in the Python programming language. This data is employed to conduct an in-depth analysis both at the national level and from a territorial perspective, examining the types and quality of Romania’s tourism supply. Thus, the analysis focuses on the number and types of accommodation establishments, the average price per room for a one-night stay during the peak summer season, as well as on the reviews provided by tourists for the accommodation units. Based on the results of this analysis, a set of recommendations is formulated to support the enhancement of Romania's tourism supply. Authorities in less developed tourist areas can boost investment through incentives and promote diverse, authentic accommodations and tourism types, while improving visibility and modernizing existing facilities.



Analysis of the data obtained through web scraping

From the analysis conducted on the data obtained through web scraping, regarding the tourism offer in Romania, it resulted that apartments represent more than half (53.38%) of the total accommodation units analyzed, indicating a clear preference for flexible, modern, and intimate lodging, especially among tourists who desire independence during their stay. The accommodation offer is diverse in terms of prices: hostels and motels are the most affordable (under 200 RON/night), preferred by budget-conscious travelers, while at the other end of the spectrum, private villas and mansions exceed 1,100 RON/night, indicating premium, exclusive options. The ratings given by tourists to accommodation establishments reflect high overall satisfaction, with tourists particularly appreciating safety, cleanliness, and the attitude of the staff (associated with scores above 9), and less so the facilities (scores around 8.8). This suggests an investment opportunity for improving equipment and facilities offered. Additionally, tourists most frequently seek facilities adapted to modern needs, such as Wi-Fi, air conditioning, parking, and pet-friendly options. Regarding the analysis of the tourism offer at the territorial level, the counties of Sibiu, Brasov, Prahova, Cluj, Timis, Constanta, as well as the Municipality of Bucharest, present the highest number of accommodation establishments, and these are also the areas where tourists have provided the most reviews for the accommodation establishments, as well as the areas with the highest average price for accommodation services (Brasov, Prahova, and Constanta). In these counties, tourists interact the most with online booking sites, constituting the most popular tourist destinations. Seven counties in the south of Romania and two in Moldova have a very low number of accommodation establishments (under 38), and in these areas, tourists have left a lower number of reviews.

Importance and implications of the research results

The analyses conducted highlight that data obtained through web scraping can provide unique perspectives on the quality of tourism offers and consumer preferences. This information is crucial for public authorities, tourism operators, and researchers, aiding in informed decision-making and the development of effective strategies to increase the competitiveness and attractiveness of tourist destinations in Romania. Thus, in counties with a low number of accommodation establishments, local authorities can stimulate investment by offering fiscal incentives, grants, or support programs for the development of guesthouses, bed-and-breakfasts, or campsites. They can also promote rural tourism, eco-tourism, and cultural tourism projects to attract investment and capitalize on local heritage. Furthermore, the lower rating given to the facilities of accommodation units indicates the need for modernization, for which co-financing or favorable credit programs can be launched for upgrading accommodation spaces. Apartments dominate the market, but authorities can encourage the development of alternative units (glamping, traditional cottages, agritourism, etc.) to offer authentic experiences and attract various tourist segments. Areas with few reviews can be integrated into digital promotion campaigns, collaborations with tourism influencers, actively listed on booking platforms, or can have local events, festivals, and thematic routes developed to increase visibility and attractiveness. This study can be expanded by using web scraping on other online booking platforms to gain an even more comprehensive view of the tourism sector. Additionally, integrating other data sources, such as social media and reviews on travel platforms, would be useful for analyzing tourist perceptions in real-time. In conclusion, using web scraping to collect and analyze data from the tourism sector provides a detailed and up-to-date picture of the market, complementing and enhancing official statistics.