

Understanding Tourism Demand: A Tree-Based Analysis of Territorial Drivers

Giulia Contu, Marco Ortu, Andrea Carta, Francesca Atzori
University of Cagliari
Department in Business and Economics

Abstract

We investigate the territorial drivers of tourism arrivals to support more effective destination planning. The analysis uses a Multivariate Regression Tree (MRT), an interpretable method that captures non-linearities and interactions while identifying homogeneous groups of destinations. The study focuses on 307 municipalities in Sardinia (Italy), using tourist arrivals by nationality as response variables and geographic, demographic, and accessibility factors as predictors. Results show that accommodation density is the main determinant of tourism attractiveness and reveal a marked coastal-inland divide driven by the interaction between supply and accessibility. The study highlights the value of tree-based methods for analyzing tourism demand in heterogeneous contexts.

Keywords: Multivariate Regression Tree; Tourism Arrivals; Destination Attractiveness; Sardinia

1 Introduction

Tourism is a key driver of regional development, and understanding what shapes tourism flows is essential for sustainable destination planning. Tourist arrivals are influenced by a wide range of factors, including accessibility, infrastructure, accommodation capacity, and local attractions [1, 2].

In this study, we explore how local structural characteristics influence tourism demand using a Multivariate Regression Tree (MRT)[3]. This method allows to jointly analyze multiple tourism markets (by nationality) while identifying groups of municipalities with similar tourism patterns. The empirical application focuses on 307 municipalities in Sardinia.

We consider tourist arrivals by nationality as multiple response variables and model them using territorial indicators such as population, transport accessibility, and accommodation supply. This approach enables us to uncover non-linear relationships and interactions between factors, as well as structural differences between coastal and inland areas.

2 Methodology

The analysis is conducted using the Multivariate Regression Tree (MRT). This method was introduced by [3] to model, explore, and describe the relationship between multivariate response data and a set of explanatory variables.

The MRT recursively partitions the data into homogeneous groups by minimizing a measure of within-group heterogeneity (impurity). Among the impurity measures proposed by [3], we focus on the *sum of squared distances*. Under this criterion, the algorithm selects splits that minimize the total within-node variance, computed as the sum of squared deviations of observations from the node centroid. It can be expressed as:

$$\sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2 \quad (1)$$

where x_{ij} denotes the value of the j -th response variable for observation i , and $\bar{x}_j$ is the mean of the j -th variable within the node. In the context of this study, MRT is used to identify groups of municipalities with similar tourism demand patterns (by nationality) and to uncover the territorial characteristics that drive these differences.

3 Results

The Multivariate Regression Tree identifies the total number of tourism facilities as the primary splitting variable, separating municipalities with low tourism capacity from those with more developed accommodation systems (Figure 1). A higher number of facilities is associated with greater tourist arrivals and higher destination attractiveness.

Transport accessibility also emerges as a relevant factor. In particular, the presence of railway stations and bus stops contributes to further splits, highlighting the role of infrastructure in supporting tourism flows. Additionally, the average length of stay—the only tourism-specific variable included—introduces further differentiation, suggesting that the duration of visits influences the composition of tourist origins and behavioral patterns.

From a spatial perspective, the identified clusters reveal a clear coastal–inland divide. Coastal municipalities are predominantly associated with higher tourism intensity, greater accommodation capacity, and better accessibility. In contrast, inland areas tend to belong to clusters characterized by lower tourism volumes and more limited infrastructure.

Overall, these results suggest that tourism concentration in Sardinia is strongly shaped by the interaction between accommodation supply, accessibility, and geographical location, with coastal proximity playing a central role in determining destination attractiveness.

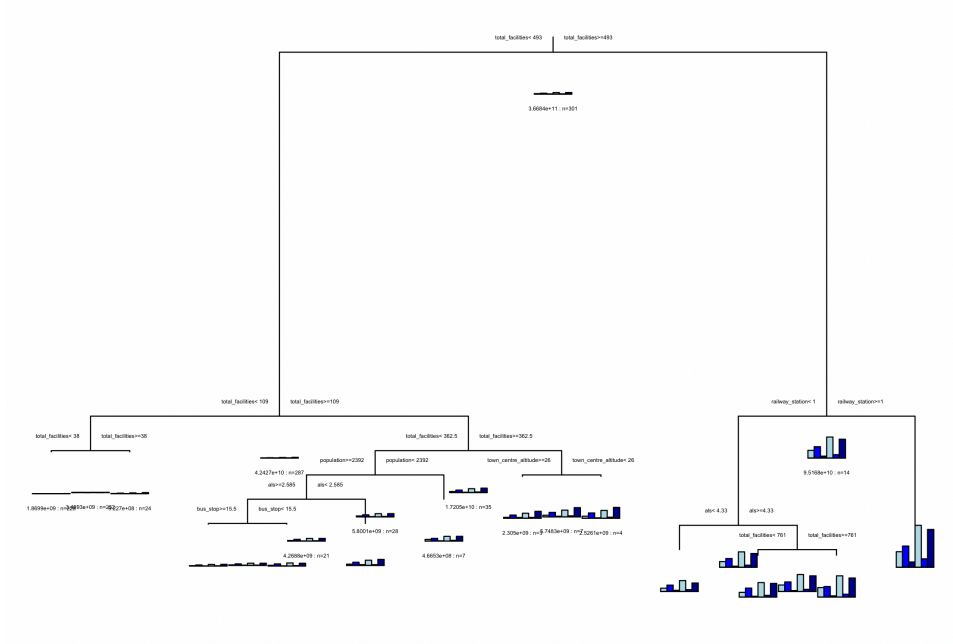


Figure 1: Multivariate tree

4 Conclusions

We jointly model multiple origin-specific demand variables and identify groups of municipalities sharing similar tourism profiles.

The results underline the key role of accommodation supply and accessibility. In particular, the number of tourism facilities emerges as the main factor distinguishing structurally peripheral municipalities from highly developed destinations. Transport infrastructure, especially railway accessibility, further contributes to differentiating areas with higher tourism demand. Moreover, the spatial distribution of clusters highlights a pronounced coastal e inland divide: coastal municipalities concentrate higher tourism intensity and infrastructure, while inland areas remain relatively less attractive.

References

- [1] Boniface, B., Cooper, C., and Cooper, R. (2016). *Worldwide Destinations: The Geography of Travel and Tourism*. Routledge.
- [2] Culiuc, A. (2014). *Determinants of International Tourism*. International Monetary Fund.
- [3] De’Ath, G. (2002). Multivariate regression trees: A new technique for modeling species–environment relationships. *Ecology*, 83(4), 1105–1117.