



UNIVERSITA' POLITECNICA DELLE MARCHE

FACULTY OF ECONOMICS

Master's Degree course International Economics and Commerce

**Un'analisi econometrica degli effetti di spillover turistico e della gestione
delle destinazioni: evidenze dalle dinamiche spaziali comunali**

**An Econometric Analysis of Tourism Spillover Effects and Destination
Management: Evidence from Municipal Spatial Dynamics**

Supervisor:

Prof. Fabio Fiorillo

Candidate:

Bibi Pari Gul

ABSTRACT

This thesis investigates the link between the intensity of tourism, geographical location and the outcomes of tourist taxes at the municipal level in Italy. The study emphasizes the differences between Core, Neighbor, Second Ring and Other municipalities in terms of tourism and tourist-tax outcomes, and the spatial connectivity between municipalities in terms of tourism and fiscal outcomes. The analysis is based on a balanced panel of 2,743 Italian municipalities in the years 2017-2024, for a total of 21,944 municipality-year observations. The study is based on a combination of tourism, tourist-tax, demographic and geographical data. The municipalities are divided into four spatial groups based on their tourism intensity and their spatial proximity to Core tourism areas. Empirical analysis is performed with ANOVA and Tukey HSD tests, Global Moran's I, baseline regression model, and Core-cluster spatial association model. The results indicate significant differences between the four spatial groups. The highest tourism intensity and tourist-tax outcomes per resident and per tourist night are found in core municipalities. The spatial analysis also reveals positive spatial association between tourism intensity and the two normalized tourist-tax measures, but not for total tourist-tax revenue. The Core-cluster analysis also shows that the results of tourist-taxes in Neighbor municipalities are more closely linked to changes in their respective Core areas than to those in Second Ring municipalities. The results indicate that there is a strong relationship between the geographical context of municipalities and tourism and tourist-tax outcomes overall. The results, however, are spatial associations and not causal spillover effects.

Contents	
ABSTRACT	5
INTRODUCTION	10
1. INSTITUTIONAL CONTEXT AND LITERATURE REVIEW	14
1.1. Italian Tourism and Institutional Context	14
1.2. Literature Review	18
1.2.1. Tourism, Local Development, and Overtourism	18
1.2.2. Tourist Taxation	21
1.2.3. Municipal Fiscal Autonomy and Strategic Interaction	24
1.2.4. Spatial Spillovers and Empirical Evidence	27
2. DATA	31
2.1. Variable Definitions	31
2.2. Research Design	34
2.3. Study Area	35
2.4. Data Sources	36
2.5. Dataset Construction	36
2.6. Variable Construction	38
2.6.1. Fiscal Outcome Variables	38
2.6.2. Independent Variables	39
2.6.3. Control Variables	40
2.6.4. Spatial Classification, Core-Cluster Construction and Spatial Weights	41
2.7. Descriptive Analysis	44
2.7.1. General Summary Statistics Table	44
2.7.2. Maps (geographical distribution)	47
2.7.3. Boxplots	52
3. METHODOLOGY	57
3.1 ANOVA and Tukey HSD	57
3.2 Global Moran's I	60
3.3 Baseline Spatial-Group Model	62
3.4 Core-Cluster Spatial Association Model	65
4. EMPIRICAL RESULTS	68
4.1. ANOVA and Tukey HSD Results	68

4.2.	Global Moran's I Results	72
4.3.	Baseline Spatial-Group Results	74
4.4.	Core-Cluster Spatial Association Results	76
5.	CONCLUSIONS	79
5.1	Main Findings	79
5.2	Policy Implications	80
5.3	Limitations and Directions for Future Research	80
	Bibliography:.....	82
	APPENDIX A — R CODE.....	86
A.1	Chapter 2 — Data Preparation and Spatial Classification.....	86
A.1.1	Data Preparation and Merging	86
A.1.2	Spatial Classification: Core, Neighbor, Second Ring and Other	95
A.1.3	Extended Tourism Panel / Tourism Variable Construction	98
A.1.4	Chapter 2 - Final Balanced Panel Construction.....	100
A.2	Chapters 3–4 — Empirical Analysis	101
A.2.1	ANOVA and Tukey HSD.....	101
A.2.2	Global Moran's I	103
A.2.3	Baseline Spatial-Group Regression	105
A.2.4	Core-Cluster Regression	108

Tables

Table 2.1 Data Sources and Variables Used in the Analysis	31
Table 2.2 Descriptive Statistics of the Main Variables.....	44
Table 4.1 One-Way ANOVA Results by Spatial Group	69
Table 4.2 Tukey HSD Pairwise Comparisons	70
Table 4.3 Global Moran's I Results.....	72
Table 4.4 Baseline Spatial-Group Regression Results	74
Table 4.5 Core-Cluster Spatial Association Results	77

Figures

Figure 1: Territorial and Structural Context	47
Figure 2: Tourism Pressure and Intensity	48
Figure 3: Tourist Tax and Fiscal Outcomes.....	49
Figure 4: Tourism Composition and Structure	50
Figure 5: Tourism Pressure and Intensity	53
Figure 6: Tourist Tax and Fiscal Outcomes.....	54

INTRODUCTION

This study explores the relationship between tourism intensity, geographical location, structural municipal characteristics and tourist-tax revenue, based on a balanced panel of 21,944 municipality-year observations from 2,743 Italian municipalities in the 2017–2024 period. It also examines if there are systematic differences in the tourist-tax outcomes across the Core, Neighbor, Second Ring, and Other spatial categories. The study merges municipal tourism, fiscal, demographic and geographical data, enriching the literature concerning local public finance, the fiscal policy of the Italian municipalities and the spatial fiscal analysis. The Italian context is suitable since the country has a large tourism industry, a variety of tourist destinations and a tax system based on the municipalities. The diffusion of this tax across many Italian municipalities generates local revenues that capture the heterogeneity of the tourism pressures that affect the different areas of the Italian national territory and provides a very rich matrix for analyzing the interaction between spatial location and structural characteristics to give rise to the fiscal effects of tourism.

The tourism sector is a significant factor in the economies of OECD countries on an international scale. According to the OECD (2026), OECD countries welcomed 819.5 million international tourist arrivals in 2024, which is 8.1% more than in 2023, and tourism is a contributor of 4.0% of OECD GDP and 19.3% of OECD services exports in 2024. There is a growing fiscal recognition of the tourist sector with the proliferation of tourist taxes. While all these benefits can be achieved through tourism, they are frequently concentrated in specific areas and may also lead to external costs

that reduce the welfare benefits of tourism (Rosselló-Nadal & Sard, 2026). These costs increase on top of others if the number of visitors to a site or destination is greater than the carrying capacity of the destination, which can be physical, ecological, social, economic, psychological, or political, and thus lowers the well-being of the visitor while raising congestion, pollution, water stress, environmental degradation, and resident dissatisfaction. Local fiscal reaction to the pressures of tourism is an important area of investigation because the thresholds are destination dependent. The question then arises on how municipalities can respond fiscally, and the analysis will focus on the instrument of the tourist tax.

A tourist tax levied on non-resident overnight stays, an important local revenue source for tourism related services and infrastructure, may be the municipalities' answer. The spatial aspect of tourism activity and policy decisions is motivation for empirical analysis because it may exceed administrative boundaries. Studies have indicated that there can be spatial spillover effects from tourist taxation, meaning that policies in one municipality can have an impact on neighboring municipalities. In Italy, these interdependencies are especially notable, since municipalities are geographically close, and tourism flows run across multiple municipalities. Providing a spatial location context to the outcome of tourists' taxes can provide information on the effectiveness and equity of local tourism fiscal policy. The spatial classification used in this thesis, Core, Neighbor, Second Ring and Other, enables a detailed examination of the impact of the proximity to the main tourism centers on the fiscal revenues of the municipalities. However, there are several issues that have not yet been answered in existing literature.

Much research has been carried out at the local or coastal level or in cross sectional analyses; however, there is limited national-level evidence. There is more specific evidence on variations in outcomes across spatial categories within a national panel at the municipality level. Joint attention of panel-econometric estimation, Italian municipal imposta di soggiorno and a tiered spatial classification has been given limited consideration. Spatial lags were found to be statistically significant for Italy in these studies, which are by Alfano et al. (2022), Conti et al. (2018), and Biagi et al. (2017), but the size of the effect differs from country to country. Furthermore, national-level evidence on tourist-tax revenues within a hierarchical spatial arrangement remains limited. Empirically, neither adoption nor rate-setting improves operating results (revenue per resident) and identification issues related to yardstick mimicking and shock autocorrelation are discussed (Revelli, 2005). This analysis, therefore, investigates the performance of the Italian tourist-tax in a panel data at a municipality level from a national perspective and compares the fiscal performance of Core, Neighbor, Second Ring and Other municipalities.

In this context, the present study explores the potential of a systematic relationship between the intensity of tourism, geographical location, and municipal features, and tourist-tax outcomes at the municipal level in Italy. It focuses on if there is a systematic difference in the intensity of tourism and the outcomes of tourist taxes between municipalities in each of the Core, Neighbor, Second Ring and Other areas of the county and if there is a spatial association between tourism and fiscal patterns across the various municipalities. The analysis also examines whether tourist-tax outcomes in Neighbor and Second Ring municipalities are correlated with changes in the

corresponding tourist-tax outcomes of their Core areas, and whether this association differs between the two groups. These questions aim to examine whether tourism pressure and its financial impacts can be regarded as the result of intertwined local tourism systems rather than as a phenomenon of the local level.

The structure of this study is as follows. The first part provides the Italian tourism and institutional context and reviews the relevant literature on tourism, tourist taxation, local public finance, and spatial interactions, concluding with the research gap that motivates the empirical analysis. The second part explains the data, the study area, data sources, the construction of the dataset, the definition of the variables, spatial classification, Core-cluster construction, the construction of spatial weights, descriptive analysis. The third part introduces the methodological framework such as ANOVA and Tukey HSD, Global Moran's I test, baseline spatial-group model and Core-cluster spatial association model. The fourth part presents and explains the empirical findings in the sequence of the methodological framework. The main results and implications for municipal tourism policy and local destination management are discussed in the thesis conclusion.

1. INSTITUTIONAL CONTEXT AND LITERATURE REVIEW

This chapter provides the conceptual and contextual foundation for the thesis. It first presents the Italian tourism and institutional context, with particular attention to the territorial distribution of tourism and the municipal tourist tax system. It then reviews the relevant literature on tourism, tourist taxation, spatial spillovers, and strategic interaction, concluding with the identification of the research gap that motivates the empirical analysis.

1.1. Italian Tourism and Institutional Context

In 2024, the number of overnight stays in hotels and other establishments that should be registered for accommodation in Italy amounted to 466.2 million (I Flussi Turistici - Anno 2024, 2026). They are unevenly distributed by ISTAT macro-regions: the North-East hosts 181 million nights (38.8%), the Centre 115.2 million nights (24.7%), the North-West 79.4 million nights (17%), the South 57.6 million nights (12.4%) and the Islands 33.0 million nights (7%) (I Flussi Turistici - Anno 2024, 2026).

Additionally, the demand from tourists who come from abroad is particularly significant in the destination regions with the highest total demand: among these, in Veneto, 70.5 per cent of overnight stays in the destinations were carried out by tourists residing abroad, followed by Lazio (67.2 per cent), Lombardia (64.8 per cent) and Trentino-Alto Adige (62.0 per cent). The growth recorded in 2024 was positive in all macro-regions, with the Islands (+7.5 per cent) and the Centre (+13.2 million overnight stays) showing the highest rates, both in terms of increases and absolute numbers (I Flussi Turistici - Anno 2024, 2026). In the city, over 42.7 million overnight stays were

recorded, which represents just over 9.2 per cent of the total of the entire country and 12.3 per cent of foreign overnight stays (I Flussi Turistici - Anno 2024, 2026).

Italy's coastal municipalities make up over 38% of the total overnight stays, while a large city accounts for 100.5 million overnight stays (21.6% of the overall figure) (I Flussi Turistici - Anno 2024, 2026). This urbanization and coactualization implies the impossibility of measuring the tourism-relevant economic and fiscal phenomena at the regional level, and the need to study the same phenomena at the local administrative level. The municipality is not only for the observation of the tourism activity, but also it is the level at which the local fiscal instruments are implemented (Ministero del Turismo, 2022; Alfano et al., 2022). Therefore, the municipality is the right territorial unit of analysis for fiscal outcomes of tourism.

An essential tool used by Italian local governments to collect income from non-residents staying overnight in Italy is the *imposizione di soggiorno*, as introduced by Legislative Decree No. 23/2011 (Alfano et al., 2022). The tax is generally capped at €5 per person per night, although qualifying provincial capitals with tourist overnight stays exceeding twenty times their resident population may apply a rate of up to €10 per night (Art. 4, paras. 1 and 1-bis, Legislative Decree No. 23/2011). It is dedicated to revenues for tourism services, to preservation of cultural and environmental assets, and other local public services. Eligible municipalities, however, are limited to a certain number, which includes the provincial capitals (*capoluoghi di provincia*), municipal unions (*unioni di comuni*), the municipalities listed on the regional tourism and art-city lists, and municipalities situated on smaller islands.

The diffusion of the tax has been widespread among the Italian municipal units. By the end of 2015, almost 1,000 municipalities had adopted either the *imposta di soggiorno* or the closely related *imposta di sbarco* (landing tax). Such municipalities represented around 70 per cent of the overnight stays of all tourists in the country, which means that the tax was applied mainly in the main tourist destinations in the country (Conti et al., 2018; Ministero del Turismo, 2022). The territorial concentration of tourism together with the geographical selectivity of the *imposta di soggiorno* despite being minor in size, and the introduction of a tax levy on them, suggest a spatial pattern in fiscal results from the tax. This territorial patterning is the reason why the municipal and spatial approach was taken in the following chapters of the thesis.

Additionally, tourism is very focused on a sub-national level. The data of ISTAT confirms this: Rome records over 42.7 million overnight stays, representing approximately 9.2% of total overnight stays in the country, while the main destination area is the North-east with 181.1 million presences (38.8%), followed by the Centre with 115.2 million presences (24.7%), (I Flussi Turistici - Anno 2024, 2026). The labor market counterpart in 2024, there has been an increase in the number of workers in the activities of tourism related services of 4.8 per cent compared to 2023 (403 thousand workers) (I Flussi Turistici - Anno 2024, 2026). These numbers thus identify the structure that the local tax must assume to be able to achieve leverage: a narrow, dense tax base and externalities that are non-trivial and are clustered around the taxpayers..

The Ministry of Tourism noted that a tribute with characteristics like the tourist tax was already applied in the legal system in Italy in the years 1910–1988 (Ministero del Turismo, 2022). Almost one thousand Comuni have introduced this *imposta* (or the

landing one) in the modern format after Legislative Decree 23/2011, growing rapidly: at the end of 2015 according to the Bank of Italy's report (Conti et al., 2018), almost one thousand Comuni introduced it. In 2024, the total number of extra-hotel establishments was 232,376 (I Flussi Turistici - Anno 2024, 2026), compared to nearly 317 thousand non-professional private accommodations. These accommodations recorded over 39 million overnight stays, which, when added to the 466.2 million overnight stays recorded in hotels and extra-hotel establishments, brought total tourist overnight stays in Italy to over 505 million (I Flussi Turistici - Anno 2024, 2026), showing an increased base (ISTAT). The development, its origin, the current restoration, the quick implementation, the formal base widening make the imposta di soggiorno a structural part of the municipal revenue accounting..

The framework is based on a model that builds on municipal autonomy with three limitations: a per-night rate cap, earmarking and regional designation. The Comuni capoluogo di provincia, the Unioni di Comuni, the Comuni included in the regional lists of tourist destinations cities as decided by the regional governments and the Comuni with legal seat in smaller islands can introduce the imposta di soggiorno; the funds raised are used to support the interventions in the tourism sector, maintenance and use and recovery of cultural and environmental goods at a local level (Conti et al., 2018). The rates and exemptions thereof and the deadlines (I Flussi Turistici - Anno 2024, 2026) are among the aspects that can be refused, according to the national directives, by the Ministry of Tourism local personalization. It is noteworthy that the region matters, since only Marche, the only region, chose to declare all municipalities in the region as tourist destinations from the moment in which the national law concerning the

accommodation tax was first enacted in the country (Alfano et al., 2022). Partial piggyback: regime with eligibility filtering in a region by regions basis invites spatial-econometric identification.

1.2. Literature Review

The literature on tourism economics highlights the growing importance of tourism for local development, while also emphasizing the challenges associated with overtourism, fiscal management, and spatial interdependencies. Recent studies focus on how tourism pressures influence local public finance decisions and how municipalities respond through taxation instruments such as tourist taxes. This section reviews the key strands of literature relevant to this thesis, covering tourism dynamics, fiscal autonomy, and spatial interactions across municipalities.

1.2.1. Tourism, Local Development, and Overtourism

Tourism offers jobs, encourages innovation and provides opportunities for local entrepreneurs and enterprises. Addressing visitor demand can lead to greater operations by local businesses and increased community services and more livable, desirable places for tourism (OECD, 2026). The effects have put tourism as an instrument of tourism-led local development. However, the economic value of tourism is no guarantee of an improvement to local welfare. Tourists, business, destinations and host communities are not evenly distributed on benefits and costs. Some locations want more visitors to spend more money in their communities; others have so many visitors that it creates a significant strain on the local community (OECD, 2026).

This conflict has been described in literature as tourism being a developmental positive and yet simultaneously having several social and environmental negative effects. On the other hand, tourism can also be a driver for economic opportunity, social interaction, cultural exchange, and building a sense of interconnectedness between people. However, unsustainable tourism can create overcrowding, pressure on infrastructure and stress on local communities. It can destroy the established community, raise housing prices, drive gentrification and push out long-term tenants. It can also strain local infrastructure, exacerbate environmental stressors, increase social conflicts, reduce social cohesion and sustainability (OECD, 2026). Due to these opposing impacts, it is not possible to judge tourism performance by the size of the tourism activity. The impact of local development on residents, public services and environmental resources upon which a destination relies also needs to be assessed by quality.

Most of these negative effects can be considered as tourism externalities, since they are costs that are borne by residents, infrastructure or environmental resources that are not necessarily costs associated with tourism transactions. According to the literature on carrying capacity, the issues arising from the high intensity of tourism are congestion, decline in service quality, environmental damage and visitor discomfort (Rosselló-Nadal & Sard, 2026). The effects of crowding are not linear. The more visitors to a destination, the easier it is to generate more demand in the early stages of destination development. Overcrowding, however, can become a problem beyond a certain point, as it can decrease the attractiveness of the destination, the perceived satisfaction, and the willingness of tourists to return (Baños-Pino et al., 2024).

The non-linear relationship between tourism growth and destination welfare provides the theoretical basis for the connection between the intensity of tourism and overtourism. The carrying capacity is the number of visitors that a destination can receive without significant degradation of the site and without a decrease in the satisfaction of the visitors. The idea is multi-dimensional and cannot be defined in a single, widely accepted way. It can be analyzed from the visitors' point of view but also the residents' point of view, the latter gaining even more significance if tourists are arriving from other places as a source of disutility to the host community (Gago et al., 2009). Excessive tourism may produce anti-tourism sentiment and feelings of loss associated with changes to the local environment. Intensities of tourism can also have a direct impact on residents' quality of life, especially in tourist areas, where residents are willing to pay for congestion reduction caused by tourists (Baños-Pino et al., 2024).

From this contrast in the benefits of tourism and its external costs, Sustainable Tourism Theory has emerged. Overall, destination management must strike a balance between economic development and environmental protection, social cohesion and the well-being of residents to ensure the sustainability of tourism. There is a strong focus on the importance of tourism rebalancing, with a greater focus on benefits than costs (OECD, 2026) and on the need to enhance the contribution of the local community to tourism governance. Therefore, tourism-driven local development must be seen as a process that can be controlled and managed instead of as a natural result of the growth in the number of tourists. Public authorities should consider the economic advantages of the extra tourism visits and the environmental and social disadvantages of going beyond the maximum sustainable level (Baños-Pino et al., 2024). This balance serves

as a foundation for analyzing tourist taxation as a policy tool for tourism destination management.

1.2.2. Tourist Taxation

Literature on tourist taxation presents this as a multi-dimensional tool, which allows us to meet the fiscal requirement of tourism and to tackle its wider social and environmental impacts. Tourism creates demand for public goods and services such as roads, security, environmental amenities and maintenance of cultural and natural assets. Tourists being non-residents may use these resources without contributing to the extra costs that are caused by their usage. Tourist taxation can thus offer a source of revenue to the municipality, as well as diversifying its revenue sources; this can lead to the reduction of reliance on other taxes and possibly contribute to improving the public service. Tourist taxation may therefore represent an additional revenue source for municipalities, and can be a way of diversifying public revenue, thus reducing dependence on other taxes, and, maybe, providing funds to improve public services (Gooroochurn & Sinclair, 2005). In this regard, taxation on tourism can help to bring in funds from tourism activity into the destination without burdening the entire fiscal responsibility of the locals.

Like the Benefit Principle, which states that people who receive benefits from a public service should pay for them. The tourist tax principle is when the user of infrastructure, amenity and service are required to pay the cost of that use through a tourist tax during their stay. The arrangement also considers Tax Exporting as most tourists from abroad are not voters in the country they visit and would not necessarily

affect local fiscal decisions (Gooroochurn & Sinclair, 2005). The domestic welfare loss of taxing and the possibility of governments financing public expenditures from a group that is not normally located in the municipality are among the potential benefits of tax exporting. But an important restriction is highlighted in the Italian literature: when the group of people who receive the benefit from the public expenditure is greater than the group of people who pay the tax (Alfano et al., 2022). Tourist taxation can then cause the reallocation of costs but is not necessarily a perfect match between beneficiaries and taxpayers.

The second one comes from Pigouvian taxation. Private prices of the tourism activity may not fully incorporate the social costs of tourism, and this can lead to congestion, environmental degradation and other externalities (Gooroochurn & Sinclair, 2005). This mismatch can be corrected by implementing a corrective tourism tax, where the activity or the users who cause an unpriced adverse effect shall be charged a tax. The tax is not only to generate revenue but also to internalize external costs and promote an efficient tourism volume (Rosselló-Nadal & Sard, 2026). However, the revenue generating and corrective goals are not always the same. Tourist taxes could raise revenue without significantly impacting the number of tourists arriving, but they might not be as effective in curbing overcrowding or other externalities as they could be if the demand for international tourism is more elastic (Rosselló-Nadal & Sard, 2026). The efficiency of a Pigouvian instrument thus relies on the price elasticity of demand and on the proximity of the tax to the costs generated by tourism.

Effective tax design is pivotal to these outcomes. The Italian framework of accommodation tax is applied to non-residents who use municipal accommodation facilities, for a period exceeding a day. It is an indirect tax levied on overnight stays, generally on a per-person, per-night basis (Alfano et al., 2022). The charge is paid by guests to the accommodation provider, who pays it to the municipal authority. This can make collection easier but can also provide an avenue for tax avoidance and tax evasion practices by accommodation managers (Rosselló-Nadal & Sard, 2026). The literature available suggests that important features of the collection mechanism are the accommodation-based collection, the tax base, the link between accommodation prices and rates and the maximum rate. It does not offer adequate indication in the material contained within the excerpts to be able to set out a general system of exemptions or to create a standard set of exemption rules for destinations.

The additional dimension of the design is achieved by using earmarking which connects tourism-tax revenues with the tourism activity that drives the need for public intervention. Within the Marche structure, funds collected need to be dedicated to tourism-related public works, supporting accommodation services, maintaining and restoring cultural and environmental heritage, and general local public services (Alfano et al., 2022). In addition, transparency in the allocation process can enhance acceptance by residents and visitors, especially when funds are allocated to sustainability, externality reduction, and local communities (Rosselló-Nadal & Sard, 2026). But taxation on tourism can have a negative impact on demand, profitability, investment and competitiveness of the destination. Another problem with inefficient or inequitable taxes occurs when they are not imposed at the welfare-maximizing levels and may give

rise to demands for lower taxes, or for retaliation, from other countries (Gooroochurn & Sinclair, 2005). Tourist taxation should therefore be conceived as a policy tool which can have different impacts depending on its objective, its level, its tax base, the distribution of the tax revenue, its administrative structure and the institutional environment in which it operates.

1.2.3. Municipal Fiscal Autonomy and Strategic Interaction

The study of tourist taxes must consider the institutional framework of municipal fiscal autonomy. Local revenue is the lifeline of local government that is essential for providing public services and meeting demands created by local economic activity. The extent of the ability of municipalities to determine their revenues, however, is determined by how revenue is shared between levels of government (Cavallero & Zagler, 2025). The Italian literature shows that in Italy the municipalities, the third tier of government, have been given limited fiscal autonomy throughout the years. The national law 42/2009 implemented fiscal federalism, aimed at enhancing the tax autonomy of municipalities and reorganization of municipal income sources. These subsequent suspensions and reintroduction of local revenues, however, created uncertainty with regards to local revenues and how these revenues might be spent, and instilled opportunities for opportunistic behavior among local policymakers (Alfano et al., 2022).

The issues raised in these developments are the core issues of Fiscal Federalism Theory. The theory is about the division of expenditure responsibilities and the division

of the power to raise expenditure between different levels of government. Municipal autonomy is relevant since local authorities can be better suited to react to destination specific conditions such as tourism intensity, local infrastructural needs and environmental pressures. Decentralization, however, does not necessarily ensure sound fiscal control. Municipalities may have limited control of their revenues and may rely on decisions made by higher levels of government, which may limit their ability to make a change in fiscal policy. As the Italian accommodation-tax framework exemplifies, there is an institutional tension between national law that opened the door to the local implementation of local tourism taxes, and the broader fiscal context of shifting revenue sources and variable fiscal allocations (Alfano et al., 2022).

Municipal autonomy also opens up opportunities for competition among municipalities. Tax Competition Theory provides an understanding of how taxes can be changed by governments in a competitive environment with a mobile tax base. In the tourism sector, it is possible to travel to destinations close by the municipality's tax choice can affect not only its own municipality, but also neighboring municipalities. Tourists are therefore considered to be a mobile base for taxation, and it is noted that a change in tax rates in one jurisdiction can influence its own tourism demand – but can also influence the tourism demand of neighboring jurisdictions (Cavallero & Zagler, 2025)). The municipalities may then have to make a choice between revenue generation from visitors and keeping their tax environment as appealing as possible compared to other tourist destinations.

This strategic setting doesn't automatically lead to identical imitation. A negative association was seen in the Marche region to the probability that another municipality adopted a tourism tax, once the municipality had instituted such a tax. The logit estimates indicate that an increase in this proportion by one point lowered the likelihood of adoption by 2.435 percentage points, while the probit estimates indicate a reduction of 2.986 percentage points. This discovery implies that towns and cities that do not implement a tax may stand out as an alternative to those that do, to draw in visitors from neighboring areas that have introduced a tax. Strategic interaction takes place not in the way that policy copying would suggest, but by means of fiscal differentiation.

Yardstick Competition Theory offers an additional account of inter-municipal dependence on policies. Local citizens can make comparisons with the budgets and actions of other nearby governments, which can encourage local policymakers to look at policies implemented by other places. There are empirical reasons to believe that local politicians do not establish fiscal policies in their isolation and local policies can be influenced by neighboring policies. Yet similar fiscal outcomes can be achieved through different mechanisms, thus making estimated intergovernmental tax reactions an inconclusive test of yardstick competition, according to Revelli (2005). Although there is no yardstick competition, there can be patterns of spatially correlated cost shocks that generate patterns of strategic policy interaction (Revelli, 2005).

It is thus important to differentiate between the concepts of tax competition, yardstick competition and common spatial influences. The examples provided by the Polish municipalities show that there is an increasing similarity of the policies on

property taxes in neighboring municipalities, with tax mimicking being identified as one of the elements of this. The average change in the neighboring tax-competition index was found to be on average 0.7 percentage points for a 1-year lag in the corresponding municipal index. However, the study specifies that it remains unclear whether this imitation is due to tax competition, yardstick competition or both (Trojanek et al., 2021). The decisions made by municipalities, therefore, should be interpreted as the result of institutional autonomy and strategic interaction in a broader spatial context.

1.2.4. Spatial Spillovers and Empirical Evidence

The fiscal aspects of tourism are not taken into consideration separately, as municipalities are geographically linked, have common markets and infrastructure and share visitors. Any adjustment of a tourist tax in one municipality can affect the competitiveness of tourist destinations in neighboring municipalities and the allocation of tourists to the surrounding region. These links also establish spatial spillovers, that is when the economic or fiscal situation of one jurisdiction has an impact on another. In tourism markets, spatial dependence can occur since tourists have the possibility to choose between geographically adjacent destinations, and municipalities can react to fiscal policies of their adjacent municipalities. Spatial Interaction Theory is a useful lens through which to examine these relationships as it highlights the flow of people and economic activity between inter-connected locations.

Spatial interaction does not always involve strategic policy behavior. Regional similarity in tax choices can be due to direct competition, policy mimicry, shared

economic characteristics, or unobserved characteristics of geographically adjacent jurisdictions. Revelli considers this a key issue in empirical studies on local governments' finances. Local government fiscal variables exhibiting spatial autocorrelation can be caused by substantive “strategic interaction” but can also be due to omitted jurisdictional characteristics or to common shocks to the tax policy. (Revelli, 2005) Further, various theories, such as expenditure spillovers, fiscal competition and yardstick competition, may lead to the same short-run pattern of intergovernmental fiscal reactions (Revelli, 2005). Therefore, it is important to carefully evaluate evidence of neighboring municipalities implementing similar policies.

The spatial dependency is present and complicated in the empirical evidence presented across the world. The study of all 2,479 Polish municipalities reveals the existence of spatial property-tax mimicking and a mechanism driving convergence of characteristics of local tax policies. In most clusters, high levels of tax-rate-reduction index in one municipality were linked to high levels in neighboring municipalities, and vice versa, as were low levels (Trojanek et al., 2021). The analysis also reveals statistical significance of the relationship between the tax reductions of the neighboring municipalities and the decisions of the municipal tax policy during the analyzed period, which suggests that the fiscal policy of neighboring municipalities has influenced the fiscal policy of the analyzed municipality (Trojanek et al., 2021). The results suggest that there are neighbor effects, though the broader literature acknowledges that imitation could be due to tax competition, yardstick competition or a combination of both.

The picture of evidence from Italy is more differentiated. The analysis of the municipalities in the Marche area shows that the choice of introducing a tourist accommodation tax was significantly influenced by the strategic interaction between municipalities which were near each other. But the association was negative: Alfano et al. (2022) report marginal-effect estimates of -2.435 in the logit model and -2.986 in the probit model. Contrary to what happened in other municipalities, local authorities did not imitate other municipalities by implementing the tax but instead seemed to differentiate themselves from them to make them more attractive in the neighboring tourism submarkets (Alfano et al., 2022). The result shows that there can be fiscal differentiation induced by spatial spillovers rather than fiscal imitation.

The findings from the Italian coastal municipalities further support the idea that spatial interaction is underpinned by geographical context. In the Tyrrhenian area, visitors shifted to less taxed municipalities, while the municipalities themselves lowered their own tax rates on the discovery by the neighboring municipalities of the same. In the study, strategic complementarity in the decisions of taxation of tourism on the shores of the Tyrrhenian Sea (Cavallero & Zagler, 2025) is found. Contrastingly, it does not highlight similar behavior in the Adriatic region (Cavallero & Zagler, 2025). This is because the spatial spillovers differ, depending on the structure of local tourism markets, as well as on the degree of substitutability between destinations.

Empirical studies point to several potential mechanisms: tourist substitutions, competition for mobile tourists, fiscal differentiation, municipal imitation, and common spatial shocks. Spatial dependence should therefore be considered as a context-specific

relationship, as opposed to a universal mechanism. The studies available suggest that the neighboring municipalities could affect each other's tourism-tax decision, but the nature of this influence, as well as the direction and magnitude, varies by region and institution. This variation involves making a distinction between actual strategic interaction and spatial correlation arising from conditions or characteristics that are not seen (Revelli, 2005).

2. DATA

This chapter outlines the data used in the analysis and explains how the dataset is constructed and prepared. It defines the main variables, describes the study area and data sources, and details the procedures used to construct fiscal, tourism, and spatial indicators. The section concludes with a descriptive analysis that provides an overview of the key characteristics and distributions of the variables across municipalities.

2.1. Variable Definitions

The empirical analysis is based on balanced data at municipality level for the years 2017-2024. This database has created by merging many official Italian databases, such as tourism statistics, demographic indicators, accommodation infrastructure, and tourist tax revenues. It serves as the foundation for analyzing the concentration of tourism, spatial spillovers, and tourist taxation at the municipal level in Italy

Table 2.1 Data Sources and Variables Used in the Analysis

Data dimension	Source	Definition
Tourist arrivals and overnight stays (municipality-level)	ISTAT	Statistics on licensed accommodation establishments are published annually but generally with a publication lag of two years (e.g., the 2012 data set was published in January 2014; Biagi et al., 2017). Demand-side figures (per IRTS conventions—arrivals and nights of stay) are not always available

Data dimension	Source	Definition
		on a monthly cadence at the municipal level, motivating the use of annual aggregates (Biagi et al., 2017).
Tourist tax revenue	Italian Ministry of the Interior; IFEL-ANCI	Aggregate transfers and budget records from municipalities that have adopted the levy under D.Lgs. 23/2011. Used by both Alfano et al. for the Marche region and by Conti et al. for the national panorama (Alfano et al., 2022; Conti et al., 2018).
Population, administrative boundaries, area	ISTAT	Population register and official shapefiles used for population, the construction of per-resident indicators, and the spatial weights matrix.
Accommodation capacity (hotels and extra-hotel)	ISTAT	ISTAT accommodation register, including hotel and extra-hotel structures. Used to construct the municipal accommodation-capacity measure (Beds) and the descriptive Extra-Hotel Share.

Data dimension	Source	Definition
Foreign-tourist residency breakdown	ISTAT	Residency-coded nights and arrivals; used to compute the foreign-share measure, which documents foreign shares exceeding 60% in Veneto, Lazio, Lombardia, and Trentino-Alto Adige and 41.3% of foreign presences in the extra-hotel sector (I Flussi Turistici - Anno 2024, 2026).
Spatial distances	ISTAT shapefiles; author's GIS calculation	For the Global Moran's I analysis, Queen-contiguity relationships are used to construct the spatial weights matrix, including municipalities that share a boundary or vertex. For municipalities without a Queen-contiguous connection in the analytical sample, a nearest-neighbour fallback based on centroid distance is used. The resulting spatial weights matrix is row-standardized. The 25-km geographical criterion is applied separately to define Second Ring municipalities in the spatial classification.

2.2. Research Design

The present study uses a quantitative research technique to analyze the relationship between tourism concentration, spatial spillovers, and tourist-tax revenue at the Italian municipal level. For this goal, the empirical analysis is based on a panel and a spatial analysis, which enables unified analysis of tourism in the temporal and spatial contexts. The empirical analysis relies on 21,944 municipality-year observations from a balanced panel of 2,743 Italian municipalities saw over the years 2017-2024. The municipality is the unit of analysis, allowing the study to capture changes within municipalities over time, as well as differences between municipalities.

Moreover, the panel data approach offers a stable framework to analyze fiscal outcomes related to tourism during the study period. This study explains how the dataset for the municipalities and variables used in the analysis were constructed. It then introduces the spatial classification of municipalities, passes on to descriptive analysis of data, and finally presents the econometric approach used to analyze the relation between tourism intensity, the geographic location, and tourist-tax revenues. Municipalities are classified in space as Core, Neighbor, Second Ring, and Other.

In addition, the Core group refers to municipalities that exhibit either persistently high tourism intensity or persistently high tourism volume. The Neighbor municipalities share an administrative boundary with at least one Core municipality, and those Second Ring municipalities are found within a 25 km buffer around Core municipalities, except for Core and Neighbor municipalities. Analysis of all empirical data is performed using R statistical software environment with structured and reproducible coding. This

method will ensure that data preparation and empirical analysis are done transparently, consistently, and replicable.

2.3. Study Area

The study area of this research is Italy. The country is one of the world's top tourist destinations and has a well-established and geographically varied tourism industry that contributes significantly to work, infrastructure and municipal public finances. The analysis is done at the municipal level, which corresponds to the administrative level at which the tourism activity is seen and the decisions concerning the tourist tax are taken. This study aims to analyze the relationship between tourism intensity and fiscal outcomes over time, from 2017 to 2024.

The Italian touristic system is very heterogeneous in terms of tourist activity and tourist tax adoption. Tourism demand is concentrated in small number of municipalities, like Rome, Venice, Florence and the Amalfi Coast, and many other municipalities have lower levels of tourist demand. In Italy, this situation is also marked by a high degree of geographical and socio-economic differences, ranging from urban and rural areas, inland and coastal regions, Alpine areas, to metropolitan centers. The institutional differences in the widespread acceptance of the tourist tax among municipalities also create variation in the intensity of tourism, accommodation structure, and tourist-tax revenues. These characteristics are a proper empirical framework for the analysis of the impact of tourism concentration, spatial spillovers, and municipal tourist-tax revenues.

2.4. Data Sources

The empirical analysis relies on a panel data at the municipality level, built by merging data from different official databases at the Italian administrative level. The final dataset is an amalgamation of data on tourism activity, accommodation infrastructure, demographic data, municipal tourist tax revenues, and geographical boundaries between the years 2017 and 2024. It uses official national data sources to create an elevated level of reliability, consistency, and national coverage. Data on tourism statistics—tourist arrivals, overnight stays, tourism capacity of accommodation and resident population--were provided by the Italian National Institute of Statistics (ISTAT).

Data on municipal tourist tax revenues were obtained from the Italian Ministry of the Interior, which publishes annually data on the tourist tax collection in Italian municipalities. The spatial classification was created from the official ISTAT municipal shapefiles, as were the neighboring municipalities, and both have been used to create the maps used in the descriptive analysis. The official municipality identification code `PRO_COM_T` was used as the common identifier for all datasets and is unique and consistent across datasets and years. The integrated database was then used to draw up the variables used in the descriptive and econometrical analysis.

2.5. Dataset Construction

The analytical dataset was built by merging several official Italian administrative and statistical data into a single panel of data at the level of the municipalities that enabled descriptive, spatial and econometric analysis. These steps included harmonizing municipality identifiers, pooling the annual observation for each

municipality available across the various data sources, deriving variables and checking the consistency and integrity of the resulting panel. All data preparation and management processes were done in a structured and reproducible workflow using the R statistical software environment.

The first merged layer included 29,403 observations of municipality-years for Italy between 2017 and 2024, and 5,083 municipalities. The various data sources and datasets were combined by the official Municipality identification code (PRO_COM_T) provided by ISTAT, which is a unique and stable identifier for all data sources and years. The variable names and the municipality identifiers were harmonized before they were merged, to ensure uniformity across the study period. A zero value was not considered as missing value during the data preparation. For instance, municipalities with no tourist-tax revenue reported zero, not because they lacked information, and when reporting a small amount, the reported amount was viewed as correct and not as underreporting.

To maintain a consistent longitudinal data set, a balanced panel was created, including only municipalities with full data for each of the eight years (2017-2024). Those municipalities that did not have a consistent observation throughout the study period were dropped from the analytical sample. This creates a balanced panel of 2,743 municipalities which are seen each year from 2017 to 2024, totaling 21,944 municipality-year observations, each municipality saw eight times per year. After construction, it was set up that there had been one observation from each municipality each year over the study period and that there were no repeated municipality-year observations. This balanced panel design is used to ensure better comparability between

municipalities and to ensure a uniform base for the following descriptive and econometric analyses.

The study period was selected to provide recent municipality-level evidence on tourism activity and tourist taxation while capturing substantial temporal variation in the Italian tourism sector. The dataset spans the years preceding the COVID-19 pandemic, the pandemic period, and the later recovery phase, thereby encompassing significant changes in tourism demand and municipal tourist-tax revenues. The implications of these temporal dynamics are examined in the empirical analysis presented in the later sections of the thesis.

2.6. Variable Construction

This Section defines the variables that make up the empirical measurement framework. The four sets of variables are distinguished: Fiscal outcome variables, variables measuring the intensity and volume of tourism activity at the municipal level, control variables to absorb the structural heterogeneity of the municipalities, and spatial variables. Furthermore, the present section has a conceptual justification focus and on the role of each variable in the empirical strategy and it is meant to be read in conjunction with that table.

2.6.1. Fiscal Outcome Variables

The main fiscal outcome is Tourist Tax Revenue, measured in euros at the municipality-year level. The aggregate measure provides the baseline fiscal outcome because it represents the total tourist-tax revenue collected by each municipality and corresponds directly to the policy-relevant municipal revenue generated under the

tourist-tax framework (Alfano et al., 2022; Conti et al., 2018; Ufficio di Statistica del Ministero del Turismo, 2022).

The analysis also employs normalized fiscal measures to capture various aspects of the performance of tourist taxes. Tourist Tax Revenue per Resident is the total tourist-tax revenue divided by resident population and represents the fiscal intensity of tourist-tax revenue relative to the resident population. Tourist Tax Revenue per Night is the total tourist-tax revenue divided by the total number of overnight stays, and represents the amount of tourist-tax revenue per overnight tourism activity. The tourist tax revenue per bed is the total tourist tax revenue divided by the number of beds available and is another descriptive measure of revenue per bed. Tourist Tax Revenue is used as the dependent variable in the baseline spatial-group regression. Tourist Tax Revenue per Resident and Tourist Tax Revenue per Night are used as outcomes in the ANOVA and Tukey HSD analysis, Global Moran's I analysis, and the Core-cluster spatial association analysis. Tourist Tax Revenue per Bed is not used as an outcome in the final regression specifications but is retained as a descriptive measure.

2.6.2. Independent Variables

The study uses three measures of tourism activity: Total Nights, Total Arrivals, and Nights per Resident. Total Nights and Total Arrivals are absolute indicators of tourism activity at the municipal level, while Nights per Resident is an indicator of tourism intensity, which is the ratio of tourism activity to the resident population (Calveras & Sákovics, 2026; I Flussi Turistici - Anno 2024, 2026). Total Nights is used to establish the persistent top-5% high-volume criterion for Core municipality

classification, while Total Arrivals is retained as an additional descriptive measure of tourism volume. The main indicator of tourism intensity is nights per Resident, which is also used as a control variable in the final regression specifications. The variables are thus used in different ways in the empirical framework: Total Nights and Nights per Resident are used in the spatial classification, while Total Arrivals is used in the descriptive analysis; Nights per Resident is also used in the regression models as a control for tourism intensity.

2.6.3. Control Variables

The final regression specifications contain four time-varying control variables: Population, Beds, Nights per Resident, and Foreign Tourist Share. The controls are added to reflect the fact that there are observable differences in the size of the municipality, the number of accommodations, the intensity of tourism, and the composition of tourists. Population reflects variations in the size of the resident population, whereas Beds reflects the capacity of accommodation available in each municipality. Nights per Resident reflects the level of tourism activity per resident and has been added to reflect differences in tourism pressure. Foreign Tourist Share (FTS) is the percentage of overnight stays that can be attributed to foreign tourists, which reflects variations in the composition of tourism demand (I Flussi Turistici - Anno 2024, 2026; Alfano et al., 2022).

Extra-Hotel Share is the ratio of extra-hotel accommodation capacity to the total accommodation capacity and is kept as a descriptive indicator of accommodation structure. It is not, however, part of the final regression specifications due to significant

missingness in the analytical sample. The addition of this variable would significantly decrease the estimation sample and impact comparability between the final models.

2.6.4. Spatial Classification, Core-Cluster Construction and Spatial Weights

The spatial classification is used to represent the geographical structure for the entire period of empirical analysis (2017–2024). The 2,743 municipalities in the balanced panel are grouped into four mutually exclusive spatial categories: Core, Neighbor, Second Ring and Other. The classification is determined by tourism intensity and spatial proximity and is kept stable throughout the study period. The Core group comprises municipalities with particularly high tourism intensity or high tourism volume, or both, whereas the Neighbor and Second Ring groups represent municipalities at increasing geographical proximity to Core municipalities. Municipalities that meet none of these criteria are identified as Other.

Two tourism criteria are used to identify the Core municipalities. First, municipalities must remain within the top 5% of the distribution of tourism intensity, measured by Nights per Resident, throughout the 2017–2024 period. Second, they must remain within the top 5% of the tourism-volume distribution, measured by Total Nights, throughout the same period. Municipalities satisfying either of these two persistent criteria are classified as Core. This union identifies 162 Core municipalities, representing approximately 5.9% of the 2,743 municipalities in the balanced panel..

The remaining municipalities are classified according to their spatial proximity to Core municipalities. Neighbor municipalities are those that share an administrative boundary with at least one Core municipality. When a Neighbor municipality is

contiguous to Core municipalities belonging to different Core clusters, it is assigned to the cluster containing the closest contiguous Core municipality, based on centroid-to-centroid distance. This identifies 481 Neighbor municipalities, representing approximately 17.5% of the balanced panel. Second Ring municipalities are those whose municipal geometries fall within the 25-km geographical criterion of one or more Core municipalities but are not contiguous to a Core municipality. This identifies 1,171 Second Ring municipalities, representing approximately 42.7% of the sample. The remaining 929 municipalities, representing 33.9% of the sample, are classified as Other. These four mutually exclusive groups form the common spatial classification used in the subsequent descriptive and econometric analyses.

The 1,171 Second Ring municipalities are assigned to Core clusters after their Second Ring membership has been established using the 25-km geographical criterion. For each Second Ring municipality, qualifying Core municipalities within this criterion are identified, and the municipality is assigned to the Core cluster containing the geographically closest qualifying Core municipality, based on centroid-to-centroid distance. This nearest-Core rule does not require a contiguity chain through Neighbor municipalities. All 1,171 Second Ring municipalities therefore receive a Core-cluster identifier, with no additional centroid-based eligibility criterion beyond the 25-km geographical criterion used to establish Second Ring membership.

Core-cluster assignment does not alter the original spatial classification. Neighbor and Second Ring municipalities remain classified according to their original groups, while the 929 Other municipalities are not allocated to Core clusters. Thus, spatial

classification identifies a municipality's spatial group, whereas Core-cluster assignment identifies the Core cluster to which an eligible municipality is geographically linked.

The final Core-cluster reference structure includes 1,814 municipalities: 162 Core, 481 Neighbor and 1,171 Second Ring municipalities. These municipalities are allocated to 81 Core clusters, with each municipality receiving one Core-cluster identifier. The structure is fixed throughout 2017–2024 and corresponds to 14,512 municipality-year observations. It provides the geographical basis for constructing cluster-level measures of Core tourist-tax outcomes and analysing their relationship with tourist-tax outcomes in Neighbor and Second Ring municipalities.

Separately from the Core-cluster structure, spatial relationships among all 2,743 municipalities are represented by a Queen-contiguity spatial weights matrix for the Global Moran's I analysis. Queen contiguity identifies municipalities that share a boundary or vertex. For municipalities without a Queen-contiguous connection, a nearest-neighbor fallback based on centroid distance is used to ensure that every municipality has at least one spatial connection. The resulting spatial structure has 10,041 non-zero spatial links, with an average of 3.66 neighbors per municipality. The spatial weights are row-standardized, so the weights associated with each municipality sum to one.

2.7. Descriptive Analysis

2.7.1. General Summary Statistics Table

This section provides an overview of the principal variables used in the empirical analysis before examining their spatial distribution across Italian municipalities. Descriptive statistics are presented to summaries the central tendency and dispersion of the fiscal, tourism, demographic, and accommodation variables included in the dataset. Establishing these overall characteristics provides the descriptive foundation for the later analysis of geographical patterns and differences across the spatial groups.

Table 2.2 Descriptive Statistics of the Main Variables

Variable	N	Mean	SD	Median	Min	Q1	Q3	Max
Arrivals	21,944	39,971.85	231,630.14	7,514.50	17	2,714.00	24,966.00	10,160,897.00
Nights	21,944	137,755.18	739,867.82	21,579.50	154	7,995.00	77,621.00	42,705,319.00
Population	21,944	15,616.80	70,363.55	5,636.00	53	2,163.00	13,235.00	2,820,219.00
Beds	21,944	1,779.85	6,992.86	426.5	24	182	1,275.00	299,983.00
Arrivals per Resident	21,944	5.65	13.04	1.45	0	1	4	254.11
Nights per Resident	21,944	21.76	56.87	4.07	0.03	1	15	1,101.89
Nights per Km ²	21,944	3,802.99	13,788.53	606.96	1.06	204	2,197.00	299,886.81
Tourist Tax Revenue	21,944	157,593.47	1,529,546.26	0	0	0	27,870.00	76,573,186.13
Tourist Tax per Resident	21,944	18.14	69.54	0	0	0	4	1,927.84
Tourist Tax per Night	21,944	0.45	0.95	0	0	0	1	32.38
Tourist Tax per Bed	21,944	34.85	82.5	0	0	0	35	4,359.30
Foreign Tourist Share	21,943	0.3516	0.2379	0.3034	0.0000	0.1587	0.5083	0.9865
Extra-Hotel Share	12,010	0.4164	0.3000	0.3452	0.0000	0.1722	0.6217	1.0000

The level of tourism activity is also highly variable, with mean numbers of annual arrivals and overnight stays equal to 39,971.85 and 137,755.18 respectively, while the median numbers of annual arrivals and overnight stays are much smaller, at 7,514.50 and 21,579.50 respectively. This difference suggests a strongly right-skewed distribution of the municipalities, with a small number having a disproportionate amount of tourism flows.

The number of arrivals (10,160,897) and overnight stays (42,705,319) also show extreme outliers in the sample. This is also reflected in population and accommodation capacity where the mean numbers are 15,616.80 people and 1,779.85 beds, respectively, but the median numbers are 5,636 people and 426.50 beds, respectively. The intensity measures show an average of 5.65 arrivals and 21.76 overnight stays per resident each year per municipality, but the medians are 1.45 and 4.07, respectively, which show that most municipalities have much lower tourism intensity. There is a national spatial concentration of tourism activity, with an average of 3,802.99 nights per km² and a median of 606.96 nights per km².

The tourism tax revenue is the most skewed, with a median of zero, and a mean of €157,593.47, meaning that a large share of municipalities has either no tourist tax or exceedingly lesser tax revenue from tourism. Medians are also zero for the revenue per capita and per bed of tourist taxes, while the mean values are €18.14 and €34.85, respectively. There are significant outliers as shown by the maximum value for tourist-tax revenue (€76,573,186.13) and its per-resident equivalent (€1,927.84). The share variables, Foreign Tourist Share and Extra-Hotel Share, are continuous proportions bounded between 0 and 1 rather than binary indicators.

Foreign Tourist Share has a mean of 0.3516 and a standard deviation of 0.2379, while Extra-Hotel Share has a mean of 0.4164 and a standard deviation of 0.3000. Their quartiles also show substantial variation across municipalities, with a first quartile of 0.1587 and a third quartile of 0.5083 for Foreign Tourist Share, and a first quartile of 0.1722 and a third quartile of 0.6217 for Extra-Hotel Share. The descriptive statistics provide an overview of the data, showing that there are considerable cross-sectional

variation and significant right skewness in most of the variables, which reflects the different tourism profiles and fiscal abilities of the Italian municipalities.

Correlation Analysis

The correlation analysis looks at the linear relationships that exist among the main tourism, demographic, and tourist-tax variables and offers a preliminary evaluation of how the variables are related before the formal econometric analysis is carried out. The results reveal a very strong positive correlation between Nights per Resident and Arrivals per Resident ($r = 0.944$), which means that municipalities having a higher level of tourism as measured by nights per resident also tend to have higher arrivals per resident.

There is a moderate positive correlation between Tourist Tax Revenue and Population ($r = 0.386$), whereas the correlations with Nights per Resident and Arrivals per Resident are fairly weak ($r = 0.103$ and 0.099 , respectively). On the other hand, Tourist Tax Revenue per Resident has stronger relationships with the tourism-intensity measures, showing a correlation of 0.714 with Nights per Resident and 0.691 with Arrivals per Resident.

Altogether, the correlation results show that there are strong relationships between the two tourism-intensity measures. The high degree of correlation between Nights per Resident and Arrivals per Resident is important for the econometric analysis since including both variables at the same time might give rise to a possible multicollinearity issue. Hence, the correlation analysis offers initial guidance regarding the choice and specification of the variables to be used in the later econometric models.

Maps 1: Spatial Classification of Italian Municipalities

The map shows the spatial classification of Italian municipalities into Core, Neighbor, Second Ring and Other clusters. The red and orange municipalities are the Core and the Neighbor municipalities, which are concentrated in the established tourism areas along the Alpine arc, along coastal corridors and in major metropolitan areas, while the yellow and grey municipalities are mostly the Second Ring and Other municipalities, located in peripheral and inland areas, with a very strong concentrated spatial structure of tourism activity.

2.7.2. Maps (geographical distribution)

This section presents the geographical distribution of key tourism and fiscal variables across Italian municipalities. The maps are organized into four thematic groups: territorial structure, tourism intensity, fiscal outcomes, and tourism composition. Together, they provide a visual overview of spatial patterns, highlighting the concentration of tourism activity, the distribution of accommodation capacity, and the variation in tourist-tax outcomes across different types of municipalities. These descriptive insights serve as a foundation for the empirical analysis developed in the following sections.

Figure 1: Territorial and Structural Context

The spatial structure and accommodation capacity of Italian municipalities are shown in figure 1. The spatial classification shows the clear spatial concentration of Core municipalities especially in major tourist areas which appear along the coastal and northern regions, surrounded by clusters of Neighbor and Second Ring municipalities. This is reflected in the distribution of accommodation beds, which also shows a great

concentration in and around these core destinations and a large proportion of municipalities have comparatively low capacity.

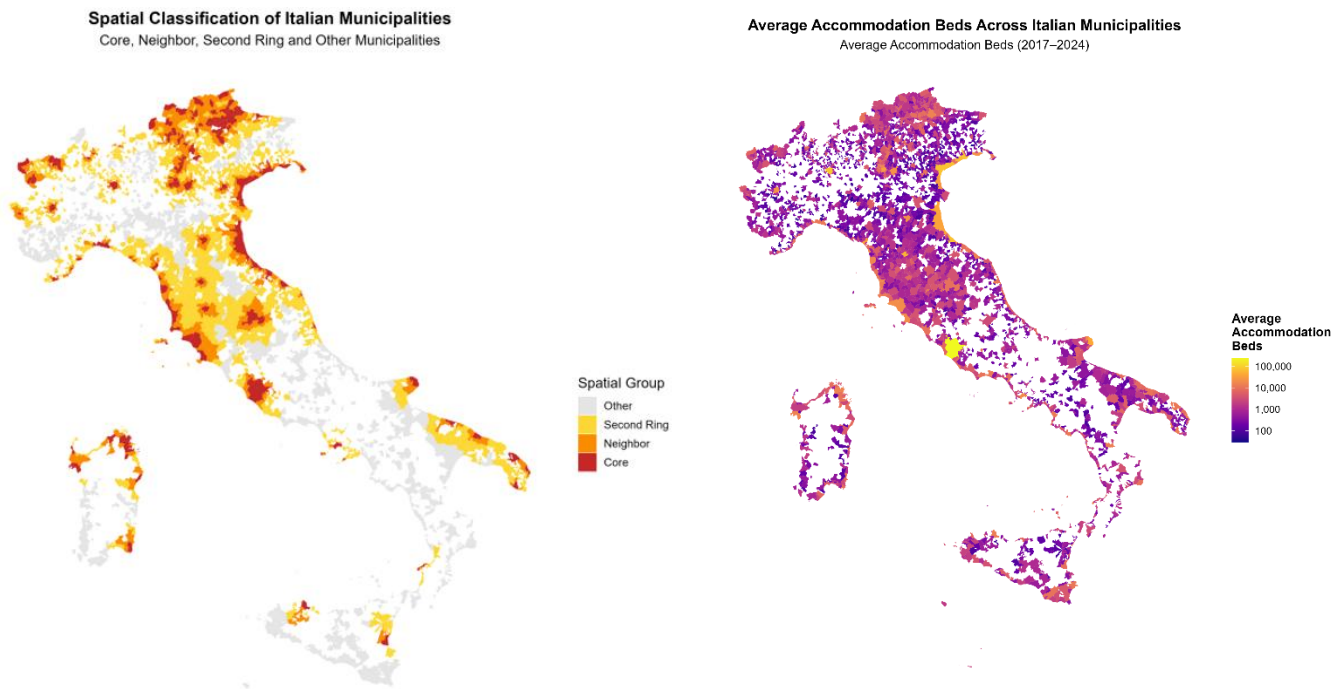


Figure 2: Tourism Pressure and Intensity

Figure 2 presents the spatial distribution of tourism pressure across Italian municipalities using three complementary indicators: total overnight stays (left), tourism intensity measured as nights per resident (centre), and total tourist arrivals (right). Together, these maps highlight the strong geographical concentration of tourism activity, particularly in northern Italy and along major coastal areas. High levels of overnight stays and arrivals are clearly clustered in well-established destinations, reflecting the uneven distribution of tourism demand across the country.

While total overnight stays and arrivals capture the absolute scale of tourism, the tourism intensity map reveals a different dimension by accounting for the size of local populations. Several smaller municipalities, especially in coastal and alpine areas, exhibit very high tourism intensity despite not having the highest absolute volumes. This suggests that tourism pressure is not only concentrated in large destinations but can also be particularly significant in smaller areas, where tourism activity represents a substantial burden relative to local capacity. These patterns provide important context for understanding spatial disparities in tourism pressure and their potential implications for local fiscal outcomes and policy responses.

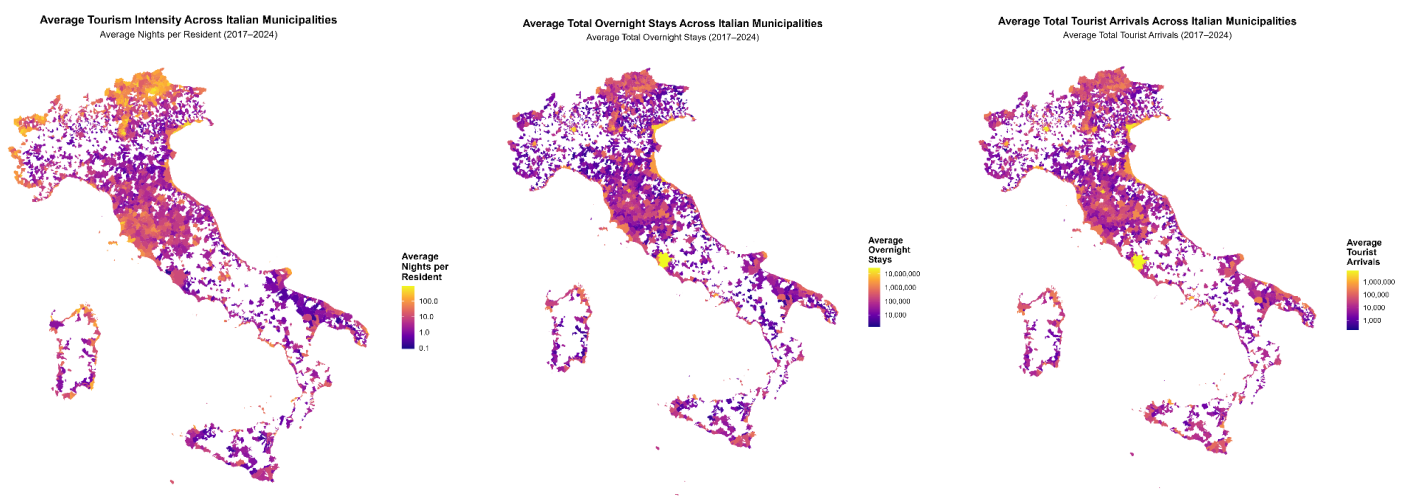


Figure 3: Tourist Tax and Fiscal Outcomes

Figure 3 illustrates the spatial distribution of tourist tax revenues across Italian municipalities using three complementary measures: total tourist tax revenue (left), tourist tax revenue per resident (center), and tourist tax revenue per overnight stay

(right). The maps reveal a strong concentration of fiscal revenues in major tourism destinations, particularly in coastal regions and established tourist hubs. Municipalities with high tourism activity tend to generate the highest levels of total tourist tax revenue, reflecting the direct link between tourism flows and local fiscal capacity.

However, when considering normalized measures, important differences appear. Tourist tax revenue per resident highlights the relative fiscal importance of tourism for local populations, with smaller municipalities often showing high values despite lower total revenues. Similarly, revenue per overnight stay provides insight into the intensity of taxation compared to tourism activity, showing variation in local tax policies and pricing structures across municipalities. Together, these indicators underline the heterogeneity of fiscal outcomes and provide a more nuanced understanding of how tourism translates into local public revenues across different spatial contexts.

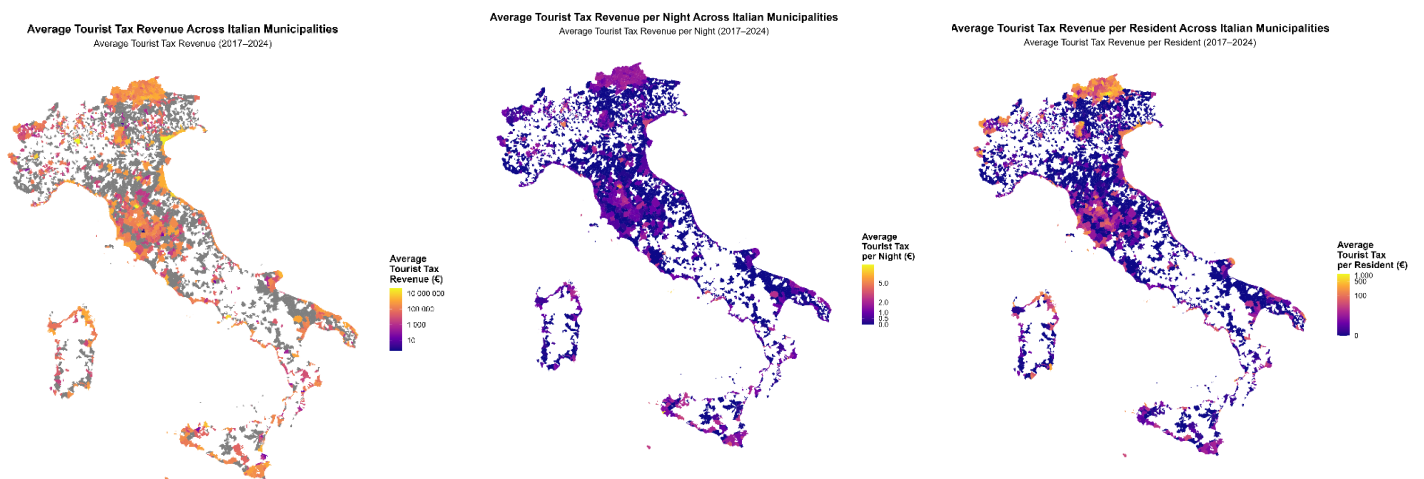
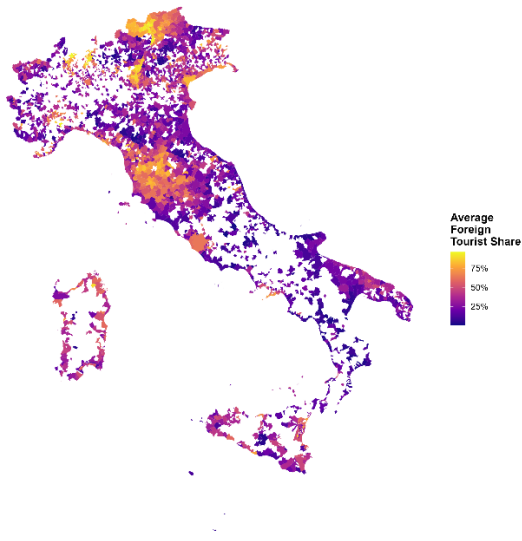


Figure 4: Tourism Composition and Structure

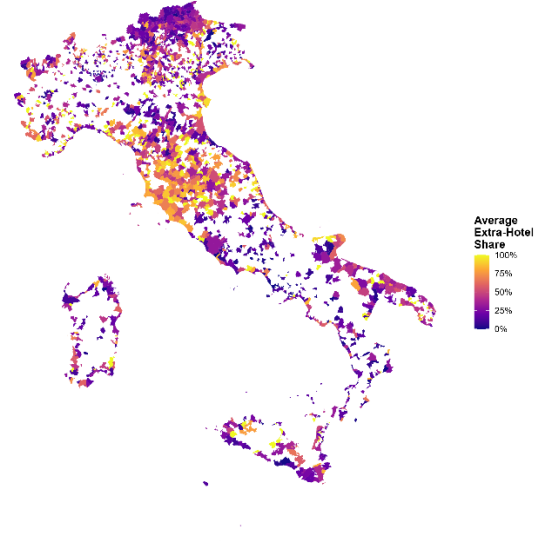
Figure 4 illustrates the composition and structural characteristics of tourism across Italian municipalities using two key indicators: the share of foreign tourists (left) and the share of extra-hotel accommodation (right). The spatial distribution of foreign tourist share reveals clear geographical patterns, with higher concentrations seen in internationally recognized destinations, particularly in northern regions and major coastal areas. In contrast, many inland and less accessible municipalities display lower shares, suggesting a stronger reliance on domestic tourism.

The extra hotel accommodation share offers additional insight into the type of tourism infrastructure across municipalities. Higher values are seen in areas where alternative forms of accommodation, such as short-term rentals, agritourism, and non-hotel lodging, are more prevalent. The comparison between the two maps highlights the heterogeneity in tourism structures, showing that municipalities differ not only in the volume of tourism but also in their composition and organization. These differences are relevant for understanding local tourism models and their potential implications for fiscal outcomes and policy design.

Average Foreign Tourist Share Across Italian Municipalities
Average Foreign Tourist Share (2017–2024)



Average Extra-Hotel Accommodation Share Across Italian Municipalities
Average Extra-Hotel Accommodation Share (2017–2024)



2.7.3. Boxplots

Tourism Pressure and Intensity To explore how tourism-related variables differ across spatial groups, boxplots are used to visualize their distributions. Boxplots provide a concise summary of the data by displaying the median, interquartile range, and potential outliers for each group. This allows for a straightforward comparison of both central tendencies and variability across locations. In the figures that follow, each box represents the distribution of a given variable within a spatial category (Core, Neighbor, Second Ring, and Other).

Moreover, the horizontal line inside each box indicates the median, while the box itself captures the middle 50% of observations. Whiskers extend to show the overall spread of the data, and individual points highlight extreme values. This approach is particularly useful for identifying spatial patterns, such as whether certain tourism characteristics are more concentrated in central areas or more dispersed across

peripheral regions. It also helps reveal the presence of skewness and heterogeneity within groups, which are important considerations for the subsequent empirical analysis.

Figure 5: Tourism Pressure and Intensity

Figure 5 illustrates the distribution of tourism pressure across spatial groups, using three key indicators: Nights per Resident (left), Total Overnight Stays (center), and Total Tourist Arrivals (right). A clear gradient emerges across all panels, with Core municipalities consistently exhibiting substantially higher values compared to Neighbor, Second Ring, and Other areas. In particular, the left panel shows that tourism intensity is significantly concentrated in Core destinations, where median values and dispersion are markedly higher, reflecting strong tourism pressure relative to local population size. Neighbor municipalities display moderate levels, while Second Ring and Other areas remain relatively low and homogeneous.

The center and right panels further reinforce this pattern in terms of absolute tourism demand. Core municipalities dominate both overnight stays and tourist arrivals, with a wide distribution and numerous extreme values, indicating the presence of major tourism hubs. In contrast, the remaining spatial groups exhibit much lower medians and limited variability, suggesting a more marginal role in the tourism system. Overall, the figure provides strong descriptive evidence of spatial concentration and highlights a clear gradient in tourism activity from Core areas toward neighboring and more peripheral municipalities.

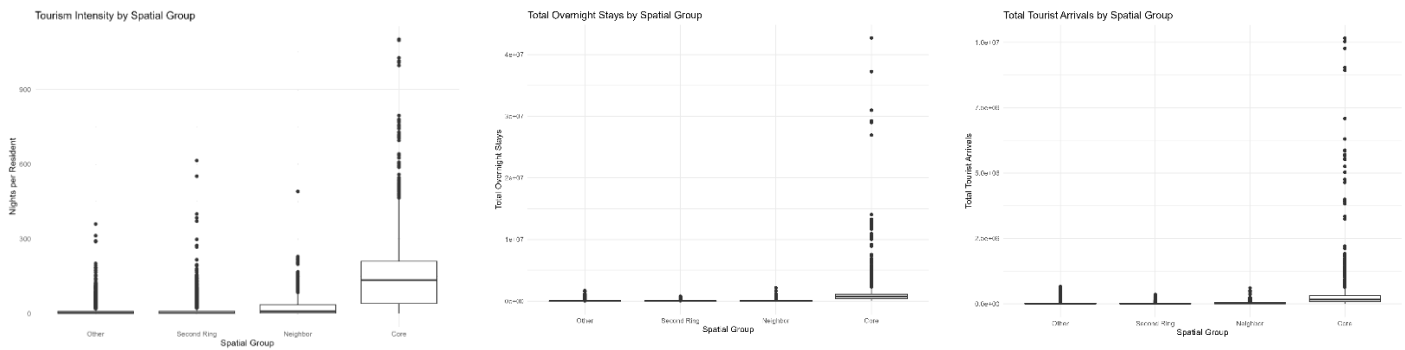


Figure 6: Tourist Tax and Fiscal Outcomes

Figure 6 presents the distribution of tourist tax revenues across spatial groups using two indicators: tourist tax revenue per resident (left) and tourist tax revenue per overnight stay (right). A clear spatial pattern emerges, with Core municipalities showing substantially higher values compared to all other groups. In the left panel, the distribution of tax revenue per resident is highly concentrated in Core destinations, with higher median values and significant dispersion, indicating that tourism contributes more intensively to local public finances in these areas. Neighbor municipalities show moderate levels, while Second Ring and Other groups remain relatively low and compressed.

The right panel further highlights differences in fiscal intensity compared to tourism activity. Core municipalities again display higher median values, indicating higher tourist-tax revenue relative to overnight tourism activity. This measure reflects the amount of tourist-tax revenue generated per overnight stay and may be influenced by factors such as tax rates, exemptions, compliance, and the composition of stays. In contrast, the remaining spatial groups show lower and more homogeneous distributions,

indicating lower tourist-tax revenue relative to overnight tourism activity. Overall, the figure suggests that both the scale of tourist-tax revenue and revenue relative to overnight tourism activity are strongly concentrated in Core destinations, while Neighbor, Second Ring, and Other municipalities display lower values.

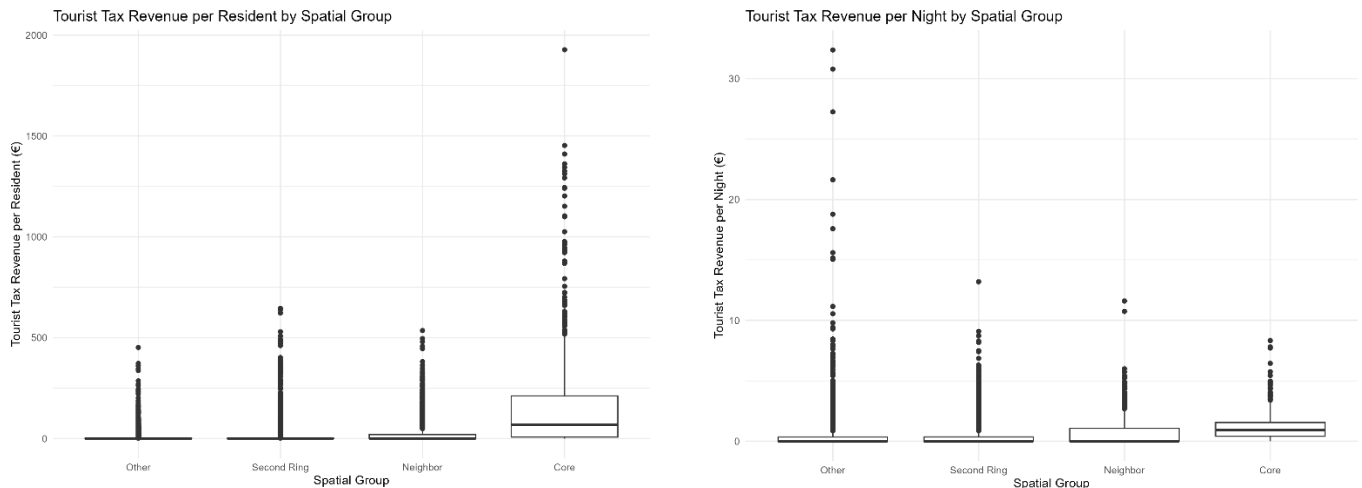
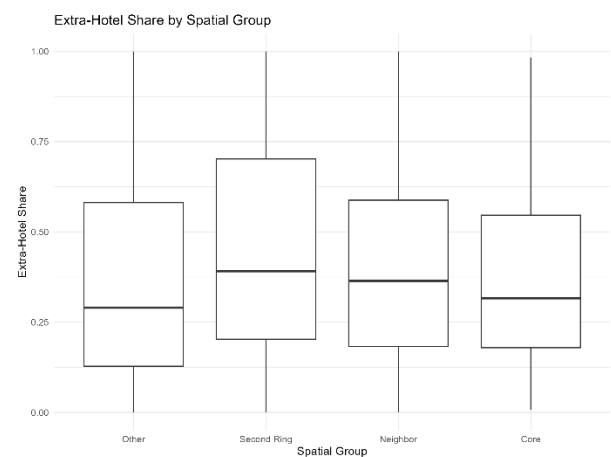
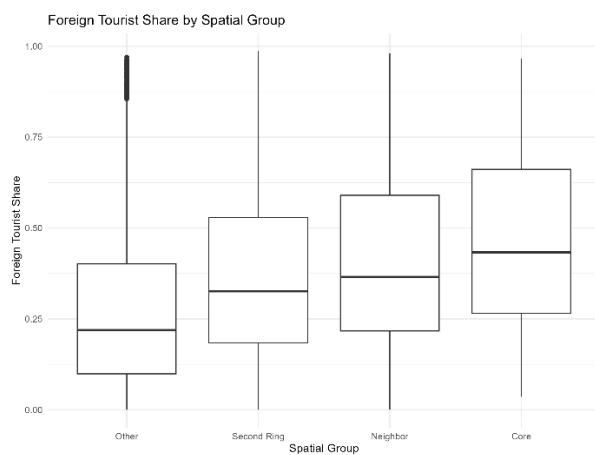


Figure 7: Tourism Composition

Figure 7 presents the distribution of tourism composition across spatial groups, using the share of foreign tourists and the share of extra-hotel accommodations. The results reveal a clear spatial gradient in the foreign tourist share. The median share increases steadily from peripheral areas (“Other”) toward the urban core. Core areas exhibit the highest concentration of international visitors, with a substantially higher median and wider dispersion compared to other groups. This indicates that central destinations are more strongly integrated into global tourism flows, while peripheral areas rely more on domestic demand.

A similar, though less pronounced, pattern emerges for the extra-hotel accommodation share. The Second Ring and Neighbor areas display relatively higher medians, suggesting a stronger presence of alternative accommodation types such as short-term rentals. In contrast, the Core shows a slightly more moderate median, indicating a more balanced mix between traditional hotels and extra-hotel options. Overall, these patterns suggest that tourism composition varies systematically across space. Central areas are characterized by a higher reliance on international tourism, while surrounding zones exhibit greater diversification in accommodation types. This spatial heterogeneity is important for understanding differences in tourism pressure and economic outcomes observed in previous sections.



3. METHODOLOGY

This chapter describes the empirical methods used to examine the relationship between tourism intensity, geographical position, municipal characteristics, and tourist-tax outcomes. The empirical strategy proceeds in four stages. First, ANOVA and Tukey HSD are used to assess differences across the four spatial groups. Second, Global Moran's I is used as a spatial diagnostic to determine whether the principal tourism and tourist-tax outcomes exhibit systematic spatial autocorrelation. Third, a baseline spatial-group model provides a conditional comparison of tourist-tax outcomes across Core, Neighbor, Second Ring, and Other municipalities. Finally, the Core-cluster spatial association model constitutes the central empirical specification and examines whether outcomes in surrounding municipalities are associated with developments in their corresponding Core tourism areas.

3.1 ANOVA and Tukey HSD

A one-way analysis of variance (ANOVA) is used to assess whether the mean values of the selected tourism-intensity variable (Nights per Resident) and tourist-tax outcomes (Tourist Tax Revenue per Resident and Tourist Tax Revenue per Night) differ across the four spatial groups: Core, Neighbor, Second Ring, and Other. The analysis is conducted separately for each variable. The results are analyzed based on the mean outcomes for each of the four categories of municipalities: Core, Neighbor, Second Ring, and Other. This is an initial statistical evaluation to determine if differences in variables of interest are related to spatial classification prior to conducting spatial autocorrelation and econometric analyses. The analysis is not used to make causal

inferences between spatial position and the outcomes analyzed; rather, it is regarded as a comparison of group means.

The unit of observation in the ANOVA is the municipality, rather than the municipality-year observation. For each municipality from the balanced panel (2,743 count) included in the 2017-2024 study period, the outcome is the average of the eight years. As such, the ANOVA dataset consists of 2,743 observations across the four spatial groups, one observation per municipality. This transformation ensures that the municipality which has been observed repeatedly each year does not get an undue influence in the group comparison, and gives a measure of the "typical" result for a municipality during the study.

Each of the three outcome variables is taken individually. Firstly, Tourist Tax Revenue per Resident determined, which captures tourist-tax revenue compared to the resident population. Secondly variable estimated is Tourist Tax Revenue per Night, which reports tourist-tax revenue to the number of tourist nights. Lastly, Nights per Resident, which defines tourism intensity relation to the resident population. For each outcome Y_i , the one-way ANOVA can be expressed as:

$$Y_i = \mu + \alpha_g + \varepsilon_i,$$

The model parameters are: Y_i is the average outcome for municipality i during the period 2017-2024; μ is the mean across the municipalities; α_g is the effect associated with the spatial group g ; ε_i is the municipality-specific error term. There are four possible values of the spatial group g : Core, Neighbor, Second Ring, and Other.

The null hypothesis for the ANOVA is that the means of the populations are the same in the four spatial groups:

$$H_0: \mu_{\text{Core}} = \mu_{\text{Neighbor}} = \mu_{\text{Second Ring}} = \mu_{\text{Other}}.$$

The alternative hypothesis is that not all of the spatial-group means are equal. The ANOVA is used to test this hypothesis using the F-statistic, which is the ratio of between-group variation to within-group variation. If the overall F-test is statistically significant, it is possible to conclude that the equality of all the means can be rejected but it is not possible to determine which groups are different. After the ANOVA, Tukey's Honestly Significant Difference (HSD) procedure is used to determine where these differences lie. The Tukey procedure will carry out all pairwise comparisons between the four spatial-group means, keeping the family-wise error rate at the level of α . There are 6 pairwise comparisons, Core–Neighbor, Core–Second Ring, Core–Other, Neighbor–Second Ring, Neighbor–Other, and Second Ring–Other, with four groups.

The estimated mean difference between two pair of groups (g, h) is

$$\hat{\Delta}_{gh} = \bar{Y}_g - \bar{Y}_h,$$

where, $\bar{Y}_g$ and $\bar{Y}_h$ represent the sample means of the groups of spatial points. The Tukey HSD procedure gives a 95% family-wise confidence interval for each pairwise difference and an adjusted p-value. The significance of the differences is thus tested with Tukey's adjusted p-values instead of pairwise tests. The procedure is carried out in R with the use of function aov to estimate the One Way ANOVA and TukeyHSD to perform the multiple comparison procedure after that. This is repeated for each of the three municipality level outcomes.

3.2 Global Moran's I

The analysis uses Global Moran's I to assess whether selected tourism and tourist-tax variables exhibit systematic spatial autocorrelation across Italian municipalities. This statistic is a measure of spatial autocorrelation and examines the tendency of municipalities with similar or dissimilar values to be grouped together. The analysis is a descriptive tool for understanding the spatial structure before the econometric models are run and is not used to establish a cause-and-effect relationship between municipalities.

The unit of observation is the municipality. The annual observations for 2017-2024 are first aggregated to give one mean value for each municipality for the study period, similar to the municipality-level approach taken in the previous group comparisons. The resulting database has 2743 municipalities, and each municipality is represented by one observation. The Global Moran's I is then determined for each of the four variables of Nights per Resident, Tourist Tax Revenue, Tourist Tax per Resident and Tourist Tax per Night.

The spatial weights matrix described in Section 2.6.4 represents the spatial relationships between municipalities. The matrix is developed, based mainly on the concept of Queen contiguity, where the municipalities are classified as neighbours if they share a boundary or a vertex with each other. The initial Queen-contiguity procedure identifies 131 municipalities with no Queen-contiguous neighbor in the analytical sample. For these municipalities, a distance-based fallback is applied by assigning the geographically nearest municipality, based on centroid-to-centroid distance, as a spatial neighbor. This ensures that every municipality has at least one

spatial connection. This produces a weights matrix that includes 2,743 spatial units and 10,041 non-zero links. The weights are row-standardized, so that the weights for the neighbors of a municipality add up to one. Global Moran's I is computed as:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

where n indicates the number of municipalities, y_i and y_j correspond to the variable for municipalities i and j , respectively, $\bar{y}$ is the sample mean of the variable, w_{ij} is the spatial weight connecting municipalities i and j , and

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$$

is the sum of all spatial weights.

Statistics quantify how much observations having similar values are clustered together. Positive Moran's I values suggest positive spatial autocorrelation, which is the tendency for municipalities with relatively high values to be near municipalities with relatively high values, and vice versa. If a value is near to the expected value in the presence of spatial randomness, then there is little evidence for systematic spatial association, while a negative Moran's I suggests a tendency for dissimilar values to be spatially adjacent. The mean value according to the randomization scheme is:

$$E(I) = -\frac{1}{n-1}.$$

Statistical significance of Moran's I is assessed using the analytical randomisation-based test implemented by the `spdep::moran.test()` function. With `zero.policy = TRUE`, the test uses the analytical expectation and variance of Moran's

I under the randomisation assumption to obtain the standardized test statistic and its corresponding p-value. The analysis uses the one-sided alternative hypothesis that Moran's I is greater than expected under spatial randomness. This procedure is analytical and does not involve Monte Carlo or permutation simulation.

The spatial diagnostic is then used as a preliminary tool for the empirical analysis, namely by using Global Moran's I. If there is significant spatial autocorrelation, then the spatial layout of the municipalities is relevant to the geographic distribution of the variable for tourism and/or tourist taxes, and it serves as motivation to take account of spatial relationships in the econometric analysis that follows. Chapter 4 provides the numerical values of Moran's I and associated test statistics and p-values, as well as interpretation.

3.3 Baseline Spatial-Group Model

After descriptive comparisons and the assessment of global spatial autocorrelation, a baseline regression model is estimated to assess whether Tourist Tax Revenue is systematically related to municipalities' spatial location with respect to Core destinations. The baseline specification includes a preliminary conditional comparison between the four spatial groups, adjusting for observable municipal characteristics and year-specific effects. It is not designed to measure the direct impact of a spillover effect or causality. The basic specification is written as:

$$Y_{it} = \alpha + \beta_1 Neighbor_i + \beta_2 SecondRing_i + \beta_3 Other_i + \gamma' X_{it} + \lambda_t + \varepsilon_{it}$$

where Y_{it} is the tourist-tax outcome for municipality i in year t , α is the constant term, $Neighbor_i$, $SecondRing_i$, and $Other_i$ are indicators for the spatial groups, X_{it} is

a vector of time-varying municipal control variables, λ_t is a vector of year fixed effects, and ε_{it} is the error term.

The Core group is excluded from the regression analysis since they are the reference group. As a result, the values for each of the three variables, β_1 , β_2 , and β_3 , represent difference values in the Neighbor, Second Ring, and Other municipalities respectively, from the Core municipalities, after controlling for the other variables and year. The spatial group indicators are time invariant, meaning that their coefficients represent the differences between the municipalities in the various spatial groups, not the changes in the relative positions of the municipalities over time.

The baseline specification's dependent variable is Tourist Tax Revenue at the municipality-year level. Population, Beds, Nights per Resident and Foreign Tourist Share are the vector of controls. Population control for differences in municipal size accounts for the capacity of accommodation available for tourism, as in the case of Beds. The Nights per Resident and Foreign Tourist Share controls the intensity of tourism demand and the composition of tourist demand, respectively. These controls are added to get conditional spatial-group comparisons instead of just unconditional differences in tourist tax revenue.

The model also add in year fixed effects, symbolized by λ_t , to account for components that affect municipalities simultaneously each year. These effects reflect the normal annual fluctuations in the tourism and fiscal context, as well as any common annual variations in the context across municipalities but vary over time. Estimates are obtained using pooled ordinary least squares (OLS) with year fixed effects for the

municipality-year observations in the balanced panel. Because the same municipalities are observed repeatedly between 2017 and 2024, regression errors may be correlated within a municipality over time. Inference therefore uses standard errors clustered at the municipality level, allowing arbitrary correlation in errors within the same municipality over time while treating different municipalities as independent clusters.

The final baseline specification does not include the Extra-Hotel Share variable. The exclusion is justified because there is substantial missing data in the panel, or about 45% of the municipality-year observations are missing. Furthermore, the missing observations are not equally distributed across spatial groups, which may lead to a significant loss of sample size and affect the comparability of the spatial-group coefficients. The final specification thus supports the complete coverage of the control variables needed for estimation of the baseline.

The baseline model is to be interpreted as a conditional association of the spatial position of a municipality with the tourist-tax revenue. The estimated coefficients show whether the municipalities in the Neighbor, Second Ring, and Other groups systematically vary from Core municipalities when the characteristics of the municipalities and common year effects are taken into account. The estimates are not intended to show that there are differences in tourist-tax revenue as a result of the spatial position, nor are the estimated spillover effects of tourism identified. Rather, the baseline model serves as a starting point for evaluating the more specific Core-cluster spatial association model.

3.4 Core-Cluster Spatial Association Model

The primary empirical specification of the thesis is the Core-cluster spatial association model. This specification investigates whether tourist-tax outcomes in Neighbor and Second Ring municipalities are systematically related to the corresponding tourist-tax outcomes of their geographically connected Core clusters. The estimation sample thus consists of only those Neighbor and Second Ring municipalities that are allocated to a Core cluster according to the procedure described in Section 2.6.4. The Core and Other municipalities are not included as the specification is about outcomes in municipalities around Core areas. The Core-cluster outcome measure $CoreOut_{ct}$, is the simple arithmetic average of the outcome measure for Core municipalities in a cluster for a given year t .

$$CoreOut_{ct} = \frac{1}{N_{ct}} \sum_{j \in C_c} Y_{jt},$$

where C_c represents Core municipalities in cluster c , Y_{jt} is the outcome for Core municipality j in year t and N_{ct} is the number of Core municipalities in cluster-year c . There is no weighting or ratio of the total of the Core municipalities. The measure is built separately for Tourist Tax Revenue per Resident and Tourist Tax Revenue per Night. The Core-cluster model is defined as:

$$Y_{it} = \alpha_i + \lambda_t + \beta CoreOut_{c(i)t} + \delta [Neighbor_i \times CoreOut_{c(i)t}] + \gamma' X_{it} + \varepsilon_{it}$$

The tourist-tax outcome of municipality i in year t is represented by Y_{it} , the outcome of the Core-cluster by $CoreOut_{ct}$ and $Neighbor_i$ is a dummy variable that equals one for Neighbor municipalities and zero for Second Ring municipalities. X_{it}

includes the time-varying controls, and α_i and λ_t are municipality and year fixed effects. Thus, Second Ring is the reference group.

The coefficient β is the association between the Core-cluster outcome and the outcome of Second Ring municipalities and the coefficient δ is the difference in this association between Neighbor municipalities. The overall association for Neighbor municipalities is thus $\beta + \delta$. The Neighbor indicator is time-invariant, and thus its direct effect is absorbed by the municipality fixed effects, but the interaction between the time-varying Core-cluster outcome and the Neighbor indicator is identified.

The model is estimated for Tourist Tax Revenue per Resident and Tourist Tax Revenue per Night. The controls are Population, Beds, Nights per Resident and Foreign Tourist Share. Extra-Hotel Share is not included due to high missingness in the estimation sample. The within estimator with municipality and year fixed effects is used to estimate the model. The estimation sample includes 1,652 municipalities (481 Neighbor and 1,171 Second Ring) and 13,216 municipality-year observations prior to the complete-case restriction. One observation Foreign Tourist Share has missing, resulting in 13,215 observations for each outcome. The same 81 Core clusters are used in the models.

Standard errors are clustered at the Core-cluster level since municipalities within the same Core cluster have the same Core-cluster outcome in a particular year. Inference is based on the CR2 cluster-robust variance estimator with Satterthwaite-adjusted degrees of freedom. The Core-cluster specification is understood as a spatial association model rather than a causal model. The coefficients show if the changes in the outcomes

of the geographically related Core clusters are systematically related to the changes in the tourist-tax outcomes for Neighbor and Second Ring municipalities, given the controls and fixed effects. The specification does not alone imply a causal tourism spillover effect.

4. EMPIRICAL RESULTS

This chapter presents the empirical findings, examining differences in tourism intensity and tourist-tax outcomes across the spatial classification of Italian municipalities. It first reports the ANOVA and Tukey HSD results, followed by the Global Moran's I spatial diagnostic. The chapter then presents the baseline spatial-group model as a preliminary conditional comparison. Finally, the Core-cluster spatial association model provides the central empirical analysis of the spatial association between surrounding municipalities and their corresponding Core tourism areas.

4.1. ANOVA and Tukey HSD Results

The first part of the empirical analysis investigates whether there are systematic differences in tourism intensity and tourist-tax outcomes among the four spatial categories (Core, Neighbor, Second Ring, Other). The analysis is carried out at the municipality level, where the average value of each variable for the period 2017–2024 is used. This data set consists of 2,743 municipalities, with one observation per municipality. Three variables are considered: Tourist Tax Revenue per Resident, Tourist Tax Revenue per Night and Nights per Resident. The first two measures reflect variations in the fiscal impact of tourism, and Nights per Resident reflects the intensity of tourism relative to residents. The ANOVA is used to determine if the means of each variable are equal among the four spatial groups.

Table 4.1 One-Way ANOVA Results by Spatial Group

Variable	df	F-statistic	p-value
Tourist Tax Revenue per Resident	3, 2739	317.40	<0.001***
Tourist Tax Revenue per Night	3, 2739	47.82	<0.001***
Nights per Resident	3, 2739	638.20	<0.001***

Notes: The ANOVA is based on 2,743 municipality-level observations, with each municipality represented by its average value over 2017–2024. The spatial groups are Core, Neighbor, Second Ring, and Other. *** $p < 0.001$.

Table 4.1 reveals that in each of three variables the null hypothesis of equal group means is rejected. The differences between the four spatial groups are statistically significant for Tourist Tax Revenue per Resident ($F = 317.40$, $p < 0.001$), Tourist Tax Revenue per Night ($F = 47.82$, $p < 0.001$) and Nights per Resident ($F = 638.20$, $p < 0.001$). The largest F-statistic is observed for Nights per Resident ($F = 638.20$), followed by Tourist Tax Revenue per Resident ($F = 317.40$) and Tourist Tax Revenue per Night ($F = 47.82$), indicating stronger between-group differentiation in tourism intensity than in the two tourist-tax measures.¹

The findings show that the intensity of tourism and the outcomes of tourist taxes are not evenly spread across the four spatial categories. In particular, they are consistent with the descriptive patterns described in Chapter 2 showing that Core municipalities had higher tourism intensity and tourist-tax outcomes than municipalities further away

¹ The particularly strong between-group differences in Nights per Resident should be interpreted in light of the spatial classification procedure, since persistent high values of Nights per Resident constitute one of the criteria used to identify Core municipalities. The corresponding ANOVA result is therefore partly related to the construction of the spatial groups.

from Core destinations. The boxplots in Figures 5 and 6 also show a clear spatial gradient within the groups in tourism pressure and fiscal outcomes, respectively.

The ANOVA F-test, however, only shows that at least one group mean is different from another group mean; it does not tell which group means differ from which other group means. Hence, the six possible pairwise comparisons are made as follows: Tukey's Honestly Significant Difference (HSD) procedure is employed to control the family-wise error rate. The results are reported in Table 4.2.

Table 4.2 Tukey HSD Pairwise Comparisons

Comparison	Tourist Tax Revenue per Resident: Difference	p-value	Tourist Tax Revenue per Night: Difference	p-value	Nights per Resident: Difference	p-value
Second Ring – Other	3.667	0.436	-0.052	0.391	1.733	0.787
Neighbor – Other	17.748	<0.001	0.170	<0.001	16.168	<0.001
Core – Other	141.335	<0.001	0.658	<0.001	151.526	<0.001
Neighbor – Second Ring	14.081	<0.001	0.222	<0.001	14.435	<0.001
Core – Second Ring	137.668	<0.001	0.711	<0.001	149.793	<0.001
Core – Neighbor	123.587	<0.001	0.489	<0.001	135.358	<0.001

Notes: Differences are calculated as the mean of the first group minus the mean of the second group. Tukey HSD uses 95% family-wise confidence intervals and adjusted p-values. *** $p < 0.001$. Results are based on 2,743 municipalities.

The Tukey results give more detailed information regarding spatial differences. The mean values for Tourist Tax Revenue per Resident are significantly higher for Core municipalities than for Other municipalities, Second Ring municipalities, and Neighbor municipalities. The mean differences for these three are, respectively, 141.335, 137.668, and 123.587, and are all statistically significant at the 1% level. Neighbor municipalities also have much higher values than Other and Second Ring municipalities. However, the difference is not statistically significant between Other and Second Ring municipalities (adjusted $p = 0.436$).

The same trend can be seen in the case of Tourist Tax Revenue per Night. The values are significantly higher in Core municipalities than in Other municipalities, Second Ring municipalities, and Neighbor municipalities, with mean difference values of 0.658, 0.711, and 0.489, respectively. Neighbor municipalities also have significantly higher values than Other and Second Ring municipalities. Again, there is no statistically significant difference in Tourist Tax Revenue per Night between Second Ring and Other municipalities (adjusted $p = 0.391$).

The highest spatial variation is exhibited by Nights per Resident. Tourism intensity is significantly higher in Core municipalities with a mean difference of 151.526 nights per resident compared to Other municipalities, and 149.793 nights per resident compared to Second Ring municipalities, and 135.358 nights per resident compared to Neighbor municipalities. Neighbor municipalities also have much higher tourism intensity than both Second Ring and Other municipalities. Tourism intensity is similar in the Second Ring and Other municipalities, with no statistically significant differences (adjusted $p = 0.787$).

The results of the ANOVA and Tukey HSD tests indicate a distinct spatial trend for tourism intensity and tourist-tax outcome. The core municipalities consistently stand apart from other groups and have significantly higher scores for all three variables. The differences involving Neighbor municipalities are also substantial, although they are generally smaller than the differences involving Core municipalities. For the three measures, no statistically significant differences are observed in Second Ring and Other municipalities. This pattern is similar to that found in Chapter 2, where the spatial classification was developed, and lends initial statistical evidence that systematic

differences in tourism pressure and tourist-tax outcomes exist with respect to proximity to Core tourism destinations.

The outcomes of the ANOVA and Tukey HSD must be interpreted as unconditional group comparisons and not as causal effects. The ANOVA and Tukey procedures do not control for other municipal characteristics or spatial dependence. They are therefore a first approximation to the description of differences between the spatial groups.

4.2. Global Moran's I Results

The main tourism and tourist-tax variables are tested for a spatial systematic relationship across Italian municipalities using Global Moran's I. The analysis is performed using municipality averages over the period 2017-2024 for 2,743 municipalities. The spatial weights matrix is row-standardized, with 10,041 non-zero links, and municipalities without a Queen-contiguity neighbor have a nearest-neighbour fallback. Table 4.3 shows the Moran's I statistics and their associated p-values estimated for the four variables examined.

Table 4.3 Global Moran's I Results

Variable	Moran's I	p-value	Statistical significance
Nights per Resident	0.135	<0.001	***
Tourist Tax Revenue	0.001	0.463	
Tourist Tax per Resident	0.198	<0.001	***
Tourist Tax per Night	0.211	<0.001	***

*Notes: Moran's I is calculated using the row-standardized spatial weights matrix. P-values are obtained from the randomisation-based Moran's I test implemented using `moran.test()` in R, with the one-sided alternative hypothesis of positive spatial autocorrelation (`alternative = "greater"`). Statistical significance: *** $p < 0.001$.*

The result exhibits a positive and statistically significant spatial autocorrelation for Nights per Resident. The Moran's I statistic is 0.135, with a p-value below 0.001,

signifying that spatially neighboring municipalities tend to exhibit similar levels of tourism intensity, rather than values being randomly distributed across the territory. A similar pattern emerges for the two normalized tourist-tax measurements. Tourist Tax per Resident exhibits a Moran's I equal to 0.198, whereas Tourist Tax per Night equals 0.211, both being statistically significant at the 0.1% level. This implies a positive spatial association in the geographical distribution of the two normalized tourist-tax measures.

On the other hand, the global spatial autocorrelation of Tourist Tax Revenue when looked at in absolute terms is not statistically significant; the Moran's I value is about 0.001 and the p-value is 0.463. As a result, the null hypothesis of spatial randomness cannot be rejected for this variable in the one-sided test conducted. Overall, the Moran's I findings indicate that tourism intensity and the normalized tourist-tax outcomes do exhibit a systematic spatial pattern among the municipalities, while absolute tourist-tax revenue fails to show a significant global spatial association. This result is in line with the use of spatial information in the later econometric analysis.

Additionally, it should be noted, though, that Global Moran's I is merely a descriptive spatial diagnostic and does not reveal any causal effects or prove the existence of tourism spillovers between municipalities. The analysis therefore moves on to the baseline spatial-group model, in which it is examined whether tourist-tax outcomes differ systematically among Core, Neighbor, Second Ring, and Other municipalities after taking into account the observed municipal characteristics and the common year effects.

4.3. Baseline Spatial-Group Results

The baseline regression gives an initial conditional comparison of tourist-tax results among the four spatial groups: Core, Neighbor, Second Ring, and Other municipalities. The year fixed effects are included and the standard errors are clustered by municipality. Core municipalities are used as the reference category, while Population, Beds, Nights per Resident and Foreign Tourist Share are control variables. The initial balanced panel contains 21,944 municipality-year observations between 2017 and 2024; one observation is dropped due to a missing value of Foreign Tourist Share, leaving an estimation sample of 21,943 observations.

Table 4.4 Baseline Spatial-Group Regression Results

Variables	Tourist Tax Revenue	Clustered SE
Constant	799,006.123	655,127.397
Other	−998,640.555*	587,220.849
Second Ring	−1,037,769.621*	586,416.080
Neighbor	−1,000,054.545*	560,743.559
2018	32,301.737***	5,137.571
2019	50,607.649***	6,381.499
2020	10,717.377	13,070.221
2021	12,493.162	12,219.585
2022	67,650.324***	7,361.000
2023	93,578.323***	16,067.619
2024	135,885.668***	19,845.775
Population	3.802	5.702
Beds	53.685	42.412
Nights per Resident	−1,902.572	1,743.334
Foreign Tourist Share	430,219.914***	127,770.110
Observations	21,943	—
R-squared	0.2145	—
Adjusted R-squared	0.2140	—

*Notes: Core municipalities are the reference spatial group, and 2017 is the reference year. Standard errors are clustered at the municipality level. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.10$. The model is estimated using pooled OLS with year fixed effects. The coefficients represent associative relationships and should not be interpreted as causal effects.*

The baseline regression results provide a conditional comparison of the performance of municipalities in the different spatial groups, using the Core

municipalities as a reference group. The Other, Second Ring, and Neighbor municipalities all have negative coefficients, with estimates of $-998,640.555$, $-1,037,769.621$, and $-1,000,054.545$, respectively. These coefficients show that the conditional Tourist Tax Revenue of the three non-Core groups is lower than Core municipalities. None of these differences are statistically significant at the 5% level, however, with p-values of 0.089, 0.077 and 0.075, respectively. Thus, at the 5% level of statistical significance, there is no statistically significant difference between the aggregate tourist-tax revenue in Core municipalities and the three non-Core groups, after controlling for municipal characteristics and common year effects.

There are also some trends in the control variables. Foreign Tourist Share is positive and statistically significant with a coefficient of 430,219.914 ($p < 0.001$). This indicates that, conditional on the spatial-group indicators, year effects, and the other controls, municipalities with a higher share of foreign tourists are associated with higher Tourist Tax Revenue. The baseline specification does not indicate that Population, Beds, and Nights per Resident are statistically significant at the 5% level. These findings should be interpreted as conditional associations and not causal effects.

The year effects are the differences in Tourist Tax Revenue across the study period compared with the reference year 2017. The 2018 and 2019 coefficients are statistically significant and positive; the 2020 and 2021 coefficients are positive but not statistically significant. The 2022, 2023 and 2024 coefficients are all positive and statistically significant, with the largest estimated coefficient for 2024. This suggests

significant temporal variation in municipal tourist-tax revenues, controlling for the spatial-group indicators and municipal controls.

Overall, the baseline results suggest that the estimated differences between the spatial groups and Core municipalities are negative but not statistically significant at the 5% level, given the municipal characteristics and year effects. The baseline specification thus provides a preliminary conditional comparison, and is not the primary test of the spatial-association argument of the thesis.

4.4. Core-Cluster Spatial Association Results

The relationship between tourist-tax outcomes in Core clusters and tourist-tax outcomes in surrounding municipalities is the main empirical analysis provided by the Core-cluster spatial association model. 1,652 Neighbor and Second Ring municipalities are analysed over 2017–2024 in 81 Core clusters, where Second Ring is the reference group. The municipality and year fixed effects and the controls Population, Beds, Nights per Resident and Foreign Tourist Share are included separately in the estimation of Tourist Tax Revenue per Resident and Tourist Tax Revenue per Night. Inference is based on CR2 cluster-robust standard errors with Satterthwaite adjusted degrees of freedom at the Core-cluster level.

Table 4.5 Core-Cluster Spatial Association Results

Variable	Tourist Tax per Resident	Tourist Tax per Night
CoreOut (β)	0.0596 $\dagger$ (0.0295)	0.0428 (0.0251)
Neighbor $\times$ CoreOut (δ)	0.0855* (0.0362)	0.1468** (0.0433)
Population	0.000519 (0.000706)	−0.0000411 (0.0000237)
Beds	0.00795 $\dagger$ (0.00379)	0.000135** (0.0000420)
Nights per Resident	0.5343 $\dagger$ (0.2309)	−0.00168 $\dagger$ (0.000642)
Foreign Tourist Share	11.12 $\dagger$ (6.24)	0.0496 (0.0931)
Year effects	Yes	Yes
Observations	13,215	13,215
Municipalities	1,652	1,652
Core clusters	81	81
Within R ²	0.243	0.123

Notes: CR2 standard errors in parentheses; Satterthwaite-adjusted degrees of freedom. $\dagger$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

The CoreOut coefficient (β) indicates the relationship between the outcomes of Core-cluster municipalities and Second Ring municipalities. It is positive in both specifications but not statistically significant at the 5% level ($p = 0.068$ in the Tourist Tax per Resident specification and $p = 0.147$ in the Tourist Tax per Night specification). Overall, there is limited evidence of an association for the Second Ring reference group. The interaction coefficient (δ) is positive and statistically significant for both outcomes: 0.0855 for Tourist Tax Revenue per Resident ($p = 0.034$) and 0.1468 for Tourist Tax Revenue per Night ($p = 0.006$). This means that the association with Core-cluster outcomes is stronger for Neighbor municipalities than for Second Ring municipalities.

The overall Neighbor association is 0.1451 for Tourist Tax Revenue per Resident with a 95% confidence interval [0.0878, 0.2024] and 0.1896 for Tourist Tax Revenue per Night with a 95% confidence interval [0.0821, 0.2971]. Both intervals do not contain zero, suggesting positive and statistically significant overall associations for Neighbor municipalities. For the controls, Beds is positively related to both outcomes, and the relationship is statistically significant at the 10% level for Tourist Tax per

Resident and the 1% level for Tourist Tax per Night. The coefficients on Population and Foreign Tourist Share are not statistically significant at the 5% level. Nights per Resident is positive and statistically significant at the 10% level in the per-resident specification, and negative and statistically significant at the 10% level in the per-night specification. The results generally show that Core-cluster tourist-tax outcomes are positively related to tourist-tax outcomes for Neighbor municipalities, and less so for the Second Ring reference group.

5. CONCLUSIONS

5.1 Main Findings

The thesis analyzed the relationship between tourism intensity, geographical location, municipal characteristics, and tourist-tax outcomes across Italian municipalities between 2017 and 2024. It was based on differences between Core, Neighbor, Second Ring and Other municipalities, spatial association between municipalities and association between Core-cluster outcomes and outcomes of surrounding municipalities. The results show systematic spatial differences and associations, and do not imply causal effects.

The first research question asked whether tourism intensity and normalized tourist-tax outcomes differ across the four spatial groups. The results show a clear spatial gradient: Core municipalities record the highest values, Neighbor municipalities generally occupy an intermediate position, while Second Ring and Other municipalities are more similar. Importantly, this pattern is observed not only for tourism intensity but also for Tourist Tax Revenue per Resident and Tourist Tax Revenue per Night.

The second research question asked whether tourism intensity and tourist-tax outcomes exhibit spatial clustering. The results show significant positive spatial autocorrelation for Nights per Resident, Tourist Tax Revenue per Resident, and Tourist Tax Revenue per Night, indicating that neighboring municipalities tend to display similar values. No significant spatial autocorrelation is found for total Tourist Tax Revenue.

The third research question was whether tourist-tax outcomes in Neighbor and Second Ring municipalities are related to outcomes in their respective Core areas, and whether this relationship differs between the two groups. The Core-cluster analysis shows a stronger association for Neighbor municipalities than for Second Ring municipalities. The overall association is positive for Neighbor municipalities for both Tourist Tax Revenue per Resident and Tourist Tax Revenue per Night, while the evidence for Second Ring municipalities is more limited.

5.2 Policy Implications

These findings suggest a role for greater coordination among municipalities, particularly where tourism activity and tourist-tax outcomes are spatially connected. The stronger association observed for Neighbor municipalities suggests that tourism-related fiscal considerations may benefit from being addressed at a broader spatial level rather than solely within individual municipal boundaries.

However, the results identify spatial associations rather than causal effects. Therefore, they do not imply that municipalities should adopt identical fiscal policies. Instead, the findings support greater consideration of the relationships between Core destinations and surrounding municipalities when designing tourism and local fiscal policies.

5.3 Limitations and Directions for Future Research

The main limitation is that the analysis identifies associations rather than causal relationships. Specifically, the Core-cluster model does not establish that changes in Core-cluster tourist-tax outcomes cause changes in tourist-tax outcomes in surrounding

municipalities. The study also uses a balanced municipal panel for 2017–2024, excluding municipalities that do not have complete observations for the entire period, and the significant missingness in Extra-Hotel Share prevents it from being included in the final regression specifications.

Future research could investigate the mechanisms underlying these spatial relationships in greater detail. It might also explore how these relationships vary in different tourism contexts and institutional settings in Italy. Overall, the findings show that tourist-tax outcomes in Italy have a meaningful spatial dimension, with the strongest relationships emerging between Core areas and their immediately neighboring municipalities.

Bibliography:

- Aguiló, E., Riera, A., & Rosselló, J. (2005). The short-term price effect of a tourist tax through a dynamic demand model. The case of the Balearic Islands. *Tourism Management*, 26(3), 359–365. <https://doi.org/10.1016/j.tourman.2003.07.005>
- Alfano, V., De Simone, E., D’Uva, M., & Gaeta, G. L. (2022). Exploring motivations behind the introduction of tourist accommodation taxes: The case of the Marche region in Italy. *Land Use Policy*, 113. <https://doi.org/10.1016/j.landusepol.2021.105903>
- Anselin, L. (2024). An Introduction to Spatial Data Science with GeoDa: Volume 1 - Exploring Spatial Data. In *An Introduction to Spatial Data Science with GeoDa: Volume 1 - Exploring Spatial Data*. CRC Press. <https://doi.org/10.1201/9781003274919>
- Baños-Pino, J. F., Boto-García, D., Zapico, E., & Mayor, M. (2024). Optimal carrying capacity in rural tourism: Crowding, quality deterioration, and productive inefficiency. *Tourism Management*, 105. <https://doi.org/10.1016/j.tourman.2024.104968>
- Biagi, B., Brandano, M. G., & Pulina, M. (2017). Tourism taxation: A synthetic control method for policy evaluation. *International Journal of Tourism Research*, 19(5), 505–514. <https://doi.org/10.1002/jtr.2123>
- Braid, R. M. (2005). Tax competition, tax exporting and higher-government choice of tax instruments for local governments. *Journal of Public Economics*, 89(9–10), 1789–1821. <https://doi.org/10.1016/j.jpubeco.2004.04.004>
- Calveras, A., & Sákovics, J. (2026). A tourist tax in a vertically segmented destination with congestion effects. *SERIEs*, 17(1), 83–105. <https://doi.org/10.1007/s13209-025-00303-2>
- Cavallero, C., & Zagler, M. (2025). The beach: tourism tax competition along the Italian coasts. *Journal of Economics/ Zeitschrift Fur Nationalökonomie*, 145(1), 77–96. <https://doi.org/10.1007/s00712-024-00890-x>

- Conti, L., Gennari, E., Quintiliani, F., Rassu, R., & Sceresini, E. (2018). L'imposta di soggiorno nei Comuni italiani. In *Ottobre*. <https://ssrn.com/abstract=3429190>
- Decreto Legislativo 14 marzo 2011, n. 23 [Legislative Decree No. 23 of 14 March 2011]. (2011). Disposizioni in materia di federalismo fiscale municipale [Provisions on municipal fiscal federalism]. Article 4.
- Elhorst, J. P. (2010). Relever le niveau de l'économetrie spatial appliquée. *Spatial Economic Analysis*, 5(1), 9–28. <https://doi.org/10.1080/17421770903541772>
- Franek, S., Wojtowicz, K., & Wyszowska, D. (2026). Tax autonomy or revenue sharing? Institutional design of personal income tax decentralization and local fiscal performance in OECD countries. *Economic Analysis and Policy*, 92, 518–544. <https://doi.org/10.1016/j.eap.2026.06.029>
- Gago, A., Labandeira, X., Picos, F., & Rodríguez, M. (2009). Specific and general taxation of tourism activities. Evidence from Spain. *Tourism Management*, 30(3), 381–392. <https://doi.org/10.1016/j.tourman.2008.08.004>
- Gooroochurn, N., & Sinclair, M. T. (2005). Economics of tourism taxation: Evidence from Mauritius. *Annals of Tourism Research*, 32(2), 478–498. <https://doi.org/10.1016/j.annals.2004.10.003>
- I flussi turistici - Anno 2024*. (2026). <https://www.istat.it/comunicato-stampa/i-flussi-turistici-anno-2024/>
- IMPOSTA DI SOGGIORNO E FLUSSI TURISTICI*. (n.d.). Retrieved July 28, 2026, from <https://www.ministeroturismo.gov.it/ufficio-di-statistica-del-ministero-del-turismo/>
- Janeba, E., & Osterloh, S. (2013). Tax and the city - A theory of local tax competition. *Journal of Public Economics*, 106, 89–100. <https://doi.org/10.1016/j.jpubeco.2013.07.004>

- Janeba, E., & Wilson, J. D. (2011). Optimal fiscal federalism in the presence of tax competition. *Journal of Public Economics*, 95(11–12), 1302–1311. <https://doi.org/10.1016/j.jpubeco.2010.11.029>
- Lee, S. K. (2014). Revisiting the impact of bed tax with spatial panel approach. *International Journal of Hospitality Management*, 41, 49–55. <https://doi.org/10.1016/j.ijhm.2014.04.010>
- Mills, B. M., Rosentraub, M. S., & Jakar, G. (2019). Tourist tax elasticity in Florida: Spatial effects of county-level room tax rate variation. *Tourism Management Perspectives*, 31, 174–183. <https://doi.org/10.1016/j.tmp.2019.05.003>
- Ministero del Turismo. (2022). Imposta di soggiorno e flussi turistici: Analisi 2019–2021. In *Ottobre*.
- Oates, W. E. (2001). Fiscal competition and European Union: contrasting q perspectives. In *Regional Science and Urban Economics* (Vol. 31). www.elsevier.nl/locate/econbase
- OECD. (2026). OECD Tourism Trends and Policies 2026. In *OECD Tourism Trends and Policies* (OECD Tourism Trends and Policies, Vol. 2026). OECD Publishing. <https://doi.org/10.1787/3FD3CD75-EN>
- Revelli, F. (2005). On Spatial Public Finance Empirics. In *International Tax and Public Finance* (Vol. 12).
- Rosselló-Nadal, J., & Sard, M. (2026). Tourism Taxation: Balancing revenues, competitiveness and sustainability in destination management. *Tourism Management*, 113. <https://doi.org/10.1016/j.tourman.2025.105326>
- Spatial Data Science*. (n.d.). Retrieved July 28, 2026, from <https://r-spatial.org/book/>
- Trojanek, M., Kisiala, W., & Trojanek, R. (2021). Do local governments follow their neighbours' tax strategies? Tax mimicking amongst Polish municipalities. *Land Use Policy*, 108. <https://doi.org/10.1016/j.landusepol.2021.105576>

- UN Tourism. (2024, November 26). *International Tourism Highlights, 2024 Edition*. International Tourism Highlights, 2024 Edition. <https://doi.org/10.18111/9789284425808>
- Wildasin, D. E., & Wilson, J. D. (n.d.). Theoretical issues in local public. economics An overview. In *Regional Science and Urban Economics* (Vol. 2).
- Wilson, J. D., & Janeba, E. (2005). Decentralization and international tax competition. *Journal of Public Economics*, 89(7), 1211–1229. <https://doi.org/10.1016/j.jpubeco.2004.08.005>
- Yang, Y., & Fik, T. (2014). Spatial effects in regional tourism growth. *Annals of Tourism Research*, 46, 144–162. <https://doi.org/10.1016/j.annals.2014.03.007>

APPENDIX A — R CODE

A.1 Chapter 2 — Data Preparation and Spatial Classification

A.1.1 Data Preparation and Merging

```
#####  
# STEP 1 - Clear Workspace  
#####  
rm(list = ls())  
cat("\014")  
#####  
# STEP 2 - Load Packages  
#####  
library(data.table)  
library(dplyr)  
#####  
# STEP 3 - Set Working Directory  
#####  
setwd("D:/Edu/Masters/masters thesis/Master_Thesis_V3")  
#####  
# STEP 4 - Define Paths  
#####  
clean_path <- "01_Data/Clean_Data"  
output_path <- "03_Output"  
#####  
# STEP 4A - Define Raw Data Path  
#####  
raw_path <- "01_Data/Raw_Data"  
#####  
# STEP 5 - Load Clean Datasets  
#####  
load(file.path(clean_path, "Tourism_Panel_2017_2024.RData"))  
load(file.path(clean_path, "Population_Panel_2017_2024.RData"))  
load(file.path(clean_path, "Tourist_Tax_Panel_2017_2024.RData"))  
load(file.path(clean_path, "Tourism_Density_2019.RData"))  
#####  
# STEP 5A - Verify Imported Datasets  
#####  
ls()  
dim(tourism_panel)  
dim(population_panel)  
dim(tourist_tax_panel)  
dim(tourism_density_panel)  
#####  
# STEP 6 - Merge Population Panel  
#####  
final_panel <- merge(  
  tourism_panel,  
  population_panel[, .(Municipality_ID, Year, Population)],  
  by = c("Municipality_ID", "Year"),  
  all.x = TRUE  
)  
#####  
# STEP 6A - Population Merge Quality Checks  
#####  
dim(final_panel)  
sum(is.na(final_panel$Population))  
summary(final_panel$Population)  
head(final_panel)  
#####  
# STEP 6B - Remaining Missing Population  
#####  
final_panel[is.na(Population), N, by = .(Municipality_ID, Municipality)][order(-N)]
```

```
#####
# STEP 7 - Separate Aggregate and Historical Municipalities
#####
missing_population <- final_panel[is.na(Population), .(Municipality_ID, Municipality, Year)]
aggregate_records <- missing_population[grepl("^Altri comuni della provincia", Municipality)]
historical_records <- missing_population[!grepl("^Altri comuni della provincia", Municipality)]
unique(historical_records$Municipality)
#####
# STEP 8 - Tourism Activity of Missing Population Records
#####
missing_population_stats <- final_panel[is.na(Population), .(
  Municipalities = uniqueN(Municipality_ID),
  Observations = .N,
  Total_Arrivals = sum(Arrivals, na.rm = TRUE),
  Total_Nights = sum(Nights, na.rm = TRUE)
)]
missing_population_stats
#####
# STEP 8A - Compare with Entire Dataset
#####
total_stats <- final_panel[, .(
  Total_Arrivals = sum(Arrivals, na.rm = TRUE),
  Total_Nights = sum(Nights, na.rm = TRUE)
)]
comparison <- data.table(
  Metric = c("Arrivals", "Nights"),
  Missing = c(missing_population_stats$Total_Arrivals, missing_population_stats$Total_Nights),
  Total = c(total_stats$Total_Arrivals, total_stats$Total_Nights)
)
comparison[, Share_Percent := round(Missing / Total * 100, 4)]
comparison
#####
# STEP 8B - Aggregate vs Historical
#####
final_panel[is.na(Population), .(
  Observations = .N,
  Arrivals = sum(Arrivals, na.rm = TRUE),
  Nights = sum(Nights, na.rm = TRUE)
), by = .(Type = ifelse(grepl("^Altri comuni della provincia", Municipality), "Aggregate", "Historical"))]
#####
# STEP 8C - Tourist Tax Revenue of Missing Population Records
#####
missing_tax <- merge(
  final_panel[is.na(Population), .(Municipality_ID, Municipality, Year)],
  tourist_tax_panel,
  by = c("Municipality_ID", "Year"),
  all.x = TRUE
)
missing_tax[, .(
  Municipalities = uniqueN(Municipality_ID),
  Observations = .N,
  Municipalities_With_Tax = uniqueN(Municipality_ID[!is.na(TouristTaxRevenue)]),
  Total_TouristTaxRevenue = sum(TouristTaxRevenue, na.rm = TRUE)
)]
#####
# STEP 8D - Compare Tourist Tax Revenue
#####
total_tax <- tourist_tax_panel[, sum(TouristTaxRevenue, na.rm = TRUE)]
comparison_tax <- data.table(
  Missing_Revenue = sum(missing_tax$TouristTaxRevenue, na.rm = TRUE),
  Total_Revenue = total_tax
)
comparison_tax[, Share_Percent := round(Missing_Revenue / Total_Revenue * 100, 4)]
comparison_tax
#####
```

```

# STEP 9 - Remove Missing Population Observations
#####
final_panel <- final_panel[!is.na(Population)]
#####
# STEP 9A - Final Panel Quality Checks
#####
dim(final_panel)
sum(is.na(final_panel$Population))
sum(duplicated(final_panel[, .(Municipality_ID, Year)]))
uniqueN(final_panel$Municipality_ID)
range(final_panel$Year)
summary(final_panel$Population)
#####
# STEP 10 - Merge Tourist Tax Panel
#####
final_panel <- merge(
  final_panel,
  tourist_tax_panel[, .(Municipality_ID, Year, TouristTaxRevenue)],
  by = c("Municipality_ID", "Year"),
  all.x = TRUE
)
#####
# STEP 10A - Tourist Tax Merge Quality Checks
#####
dim(final_panel)
sum(is.na(final_panel$TouristTaxRevenue))
summary(final_panel$TouristTaxRevenue)
uniqueN(final_panel[!is.na(TouristTaxRevenue), Municipality_ID])
sum(duplicated(final_panel[, .(Municipality_ID, Year)]))
#####
# STEP 11 - Replace Missing Tourist Tax with Zero
#####
final_panel[is.na(TouristTaxRevenue), TouristTaxRevenue := 0]
#####
# STEP 11A - Tourist Tax Final Checks
#####
sum(is.na(final_panel$TouristTaxRevenue))
summary(final_panel$TouristTaxRevenue)
table(final_panel$TouristTaxRevenue == 0)
#####
# STEP 12 - Merge Tourism Density
#####
final_panel <- merge(
  final_panel,
  tourism_density_panel[, .(Municipality_ID, Tourism_Category_Code, Tourism_Category, Density_Index, Pressure_Index,
Tourist_Index, Tourism_Density_Class)],
  by = "Municipality_ID",
  all.x = TRUE
)
#####
# STEP 12A - Tourism Density Quality Checks
#####
dim(final_panel)
sum(is.na(final_panel$Tourism_Density_Class))
uniqueN(final_panel$Municipality_ID)
sum(duplicated(final_panel[, .(Municipality_ID, Year)]))
table(final_panel$Tourism_Density_Class, useNA = "ifany")
head(final_panel)
#####
# STEP 12B - Investigate Missing Tourism Density
#####
final_panel[is.na(Tourism_Density_Class), .(Municipality_ID, Municipality, Province, Region)][order(Municipality_ID)]
#####
# STEP 12C - Check Missing Municipality IDs in Tourism Density
#####

```

```

tourism_density_panel[Municipality_ID %in% c("001029", "001045", "001075", "108049", "111028", "111079")]
#####
# STEP 13 - Replace Missing Tourism Density Class
#####
final_panel[is.na(Tourism_Density_Class), Tourism_Density_Class := "S0"]
final_panel[is.na(Density_Index), Density_Index := "D0"]
final_panel[is.na(Pressure_Index), Pressure_Index := "P0"]
final_panel[is.na(Tourist_Index), Tourist_Index := "T0"]
#####
# STEP 13A - Tourism Density Final Checks
#####
sum(is.na(final_panel$Tourism_Density_Class))
sum(is.na(final_panel$Density_Index))
sum(is.na(final_panel$Pressure_Index))
sum(is.na(final_panel$Tourist_Index))
table(final_panel$Tourism_Density_Class)
#####
# STEP 14 - Final Master Panel Audit
#####
nrow(final_panel)
uniqueN(final_panel$Municipality_ID)
min(final_panel$Year)
max(final_panel$Year)
sum(duplicated(final_panel[, .(Municipality_ID, Year)]))
sum(is.na(final_panel$Population))
sum(is.na(final_panel$TouristTaxRevenue))
sum(is.na(final_panel$Tourism_Density_Class))
#####
# STEP 14A - Missing Values
#####
data.table(
  Variable = names(final_panel),
  Missing = sapply(final_panel, function(x) sum(is.na(x)))
)
#####
# STEP 14B - Dataset Structure
#####
str(final_panel)
#####
# STEP 14C - Summary Statistics
#####
summary(final_panel)
#####
# STEP 14D - Save Master Panel
#####
save(final_panel, file = file.path(clean_path, "Master_Analysis_Panel_2017_2024.RData"))
fwrite(final_panel, file = file.path(clean_path, "Master_Analysis_Panel_2017_2024.csv"))
#####
# STEP 15 - Replace Missing Tourism Categories
#####
final_panel[is.na(Tourism_Category_Code), Tourism_Category_Code := "Q"]
final_panel[is.na(Tourism_Category), Tourism_Category := "Comuni non turistici"]
sum(is.na(final_panel$Tourism_Category_Code))
sum(is.na(final_panel$Tourism_Category))
#####
# STEP 16 - Load Municipality Registry
#####
list.files(file.path(raw_path, "Municipality_Registry"))
#####
# STEP 16A - List Municipality Registry Files
#####
list.files(file.path(raw_path, "Municipality_Registry", "Limiti01012024"))
list.files(file.path(raw_path, "Municipality_Registry", "Limiti01012024"), recursive = TRUE)
#####
# STEP 16B - Load Required Package

```

```
#####
library(sf)
#####
# STEP 16C - Read Municipality Shapefile
#####
municipality_map <- st_read(file.path(raw_path, "Municipality_Registry", "Com01012024", "Com01012024_WGS84.shp"),
quiet = TRUE)
names(municipality_map)
head(municipality_map)
st_crs(municipality_map)
#####
# STEP 16D - Create Municipality Area Dataset
#####
municipality_area <- municipality_map |> st_drop_geometry() |> as.data.table()
municipality_area <- municipality_area[, .(Municipality_ID = PRO_COM_T, Area_km2 = Shape_Area / 1000000)]
#####
# STEP 16E - Area Quality Checks
#####
dim(municipality_area)
head(municipality_area)
summary(municipality_area$Area_km2)
sum(is.na(municipality_area$Area_km2))
sum(duplicated(municipality_area$Municipality_ID))
#####
# STEP 16F - Merge Area
#####
final_panel <- merge(final_panel, municipality_area, by = "Municipality_ID", all.x = TRUE)
#####
# STEP 16G - Area Merge Quality Checks
#####
dim(final_panel)
sum(is.na(final_panel$Area_km2))
summary(final_panel$Area_km2)
sum(duplicated(final_panel[, .(Municipality_ID, Year)]))
#####
# STEP 17 - Create Pre-Period Tourism Intensity (2017-2019)
#####
pre_period_intensity <- final_panel[Year %in% 2017:2019, .(Avg_Nights = mean(Nights, na.rm = TRUE), Avg_Population
= mean(Population, na.rm = TRUE)), by = Municipality_ID]
pre_period_intensity[, PrePeriod_TourismIntensity := Avg_Nights / Avg_Population]
pre_period_intensity <- pre_period_intensity[, .(Municipality_ID, PrePeriod_TourismIntensity)]
#####
# STEP 17A - Quality Checks
#####
dim(pre_period_intensity)
summary(pre_period_intensity$PrePeriod_TourismIntensity)
sum(is.na(pre_period_intensity$PrePeriod_TourismIntensity))
head(pre_period_intensity)
#####
# STEP 17B - Merge Pre-Period Tourism Intensity
#####
final_panel <- merge(final_panel, pre_period_intensity, by = "Municipality_ID", all.x = TRUE)
#####
# STEP 17C - Final Checks
#####
sum(is.na(final_panel$PrePeriod_TourismIntensity))
summary(final_panel$PrePeriod_TourismIntensity)
head(final_panel[, .(Municipality_ID, Year, PrePeriod_TourismIntensity)])
final_panel[is.na(PrePeriod_TourismIntensity), PrePeriod_TourismIntensity := 0]
sum(is.na(final_panel$PrePeriod_TourismIntensity))
summary(final_panel$PrePeriod_TourismIntensity)
#####
# STEP 18 - Create Analysis Variables
#####
# Tourist Tax Dummy
```

```

final_panel[, TouristTaxDummy := as.integer(TouristTaxRevenue > 0)]
# Arrivals per Resident
final_panel[, ArrivalsPerResident := fifelse(Population > 0, Arrivals / Population, NA_real_)]
# Nights per Resident
final_panel[, NightsPerResident := fifelse(Population > 0, Nights / Population, NA_real_)]
# Nights per km²
final_panel[, NightsPerKm2 := fifelse(Area_km2 > 0, Nights / Area_km2, NA_real_)]
# Tourist Tax Revenue per Resident
final_panel[, TouristTaxPerResident := fifelse(Population > 0, TouristTaxRevenue / Population, NA_real_)]
# Tourist Tax Revenue per Night
final_panel[, TouristTaxPerNight := fifelse(Nights > 0, TouristTaxRevenue / Nights, 0)]
#####
# STEP 18A - Analysis Variable Quality Checks
#####
summary(final_panel[, .(TouristTaxDummy, ArrivalsPerResident, NightsPerResident, NightsPerKm2,
TouristTaxPerResident, TouristTaxPerNight)])
sapply(final_panel[, .(TouristTaxDummy, ArrivalsPerResident, NightsPerResident, NightsPerKm2, TouristTaxPerResident,
TouristTaxPerNight)], function(x) sum(is.na(x)))
#####
# STEP 19 - Save Master Analysis Panel (Version 1)
#####
save(final_panel, file = file.path(clean_path, "Master_Analysis_Panel_v1_2017_2024.RData"))
fwrite(final_panel, file = file.path(clean_path, "Master_Analysis_Panel_v1_2017_2024.csv"))
#####
# STEP 20 - Inspect Tourism Workbook
#####
library(readxl)
excel_sheets(file.path(raw_path, "Tourism", "2. Dati comunali 2014-2024.xlsx"))
#####
# STEP 20A - Inspect 2017 Sheet
#####
tourism_check <- read_excel(file.path(raw_path, "Tourism", "2. Dati comunali 2014-2024.xlsx"), sheet = "2017", col_names
= FALSE, n_max = 15)
tourism_check
tourism2017 <- read_excel(file.path(raw_path, "Tourism", "2. Dati comunali 2014-2024.xlsx"), sheet = "2017", skip = 5)
names(tourism2017)
head(tourism2017)
#####
# STEP 20 - Load Extended Tourism Panel
#####
load(file.path(clean_path, "Tourism_Panel_Extended_2017_2024.RData"))
#####
# STEP 21A - Merge Quality Checks
#####
dim(final_panel)
sum(duplicated(final_panel[, .(Municipality_ID, Year)]))
sapply(final_panel[, .(Arrivals, Nights, Population, TouristTaxRevenue)], function(x) sum(is.na(x)))
names(final_panel)
#####
# STEP 22 - Inspect Monthly Tourism Sheet
#####
monthly_check <- read_excel(file.path(raw_path, "2. Dati comunali 2014-2024.xlsx"), sheet = "Dati Mensili-Monthly data",
col_names = FALSE, n_max = 20)
#####
# STEP 22A - Sheet Structure
#####
dim(monthly_check)
names(monthly_check)
#####
# STEP 22B - Inspect Monthly Headers
#####
monthly_headers <- read_excel(
file.path(raw_path, "2. Dati comunali 2014-2024.xlsx"),
sheet = "Dati Mensili-Monthly data",
col_names = FALSE,

```

```

n_max = 6
)
monthly_headers
as.Date(c(44562,44593,44621,44652,44682), origin = "1899-12-30")
#####
# STEP 23 - Final Master Panel Audit
#####
"\n"
"FINAL MASTER PANEL"
"Observations      :", nrow(final_panel), "\n"
"Municipalities    :", uniqueN(final_panel$Municipality_ID), "\n"
"Years             :", min(final_panel$Year), "-", max(final_panel$Year), "\n"
"Variables         :", ncol(final_panel), "\n"
"Duplicate Rows    :", sum(duplicated(final_panel[, .(Municipality_ID, Year)])), "\n"
"Missing Values    :", sum(is.na(final_panel)), "\n"
#####
# STEP 23A - Missing Values
#####
data.table(
  Variable = names(final_panel),
  Missing = sapply(final_panel, function(x) sum(is.na(x)))
)
#####
# STEP 23B - Save Final Dataset
#####
final_path <- file.path("01_Data", "Final_Data")
if (!dir.exists(final_path)) {
  dir.create(final_path, recursive = TRUE)
}
save(final_panel, file = file.path(final_path, "Master_Analysis_Panel_Final_2017_2024.RData"))
fwrite(final_panel, file = file.path(final_path, "Master_Analysis_Panel_Final_2017_2024.csv"))
#####
# STEP 23 - Load Final Panel for Remaining Variables
#####
final_path <- "D:/Edu/Masters/masters thesis/Master_Thesis_V3/01_Data/Final_Data"
load(file.path(final_path, "Master_Analysis_Panel_Final_2017_2024.RData"))
#####
# STEP 23A - Load Required Packages
#####
library(data.table)
library(dplyr)
library(readxl)
library(sf)
#####
# STEP 23B - Set Project Paths
#####
raw_path <- "01_Data/Raw_Data"
clean_path <- "01_Data/Clean_Data"
final_path <- "01_Data/Final_Data"
#####
# STEP 23C - Load Final Panel
#####
load(file.path(clean_path, "Master_Analysis_Panel_Final_2017_2024.RData"))
#####
# STEP 23D - Inspect Final Panel
#####
dim(final_panel)
names(final_panel)
head(final_panel)
str(final_panel)
#####
# STEP 23E - Quality Checks
#####
sum(duplicated(final_panel[, .(Municipality_ID, Year)]))
summary(final_panel)

```

```
#####
# STEP 24A - List Accommodation Capacity Files
#####
list.files(file.path(raw_path, "Accommodation_Capacity"))
list.files(file.path(raw_path, "Accommodation_Capacity"), recursive = TRUE)
raw_path
list.files(raw_path)
list.files("D:/Masters/masters thesis/Master_Thesis_V3/01_Data/Raw_Data/Accommodation_Capacity")
list.files("D:/Masters/masters thesis/Master_Thesis_V3/01_Data/Raw_Data/Accommodation_Capacity", recursive = TRUE)
#####
# STEP 23B - Set Absolute Project Paths
#####
raw_path <- "D:/Masters/masters thesis/Master_Thesis_V3/01_Data/Raw_Data"
clean_path <- "D:/Masters/masters thesis/Master_Thesis_V3/01_Data/Clean_Data"
final_path <- "D:/Masters/masters thesis/Master_Thesis_V3/01_Data/Final_Data"
list.files(raw_path)
#####
# STEP 24B - Inspect Accommodation Workbook
#####
library(readxl)
excel_sheets(file.path(raw_path, "Accommodation_Capacity", "Capacità comunale 2013-2024.xlsx"))
#####
# STEP 24C - Inspect 2017 Accommodation Sheet
#####
raw_path <- "D:/Edu/Masters/masters thesis/Master_Thesis_V3/01_Data/Raw_Data"
clean_path <- "D:/Edu/Masters/masters thesis/Master_Thesis_V3/01_Data/Clean_Data"
final_path <- "D:/Edu/Masters/masters thesis/Master_Thesis_V3/01_Data/Final_Data"
accommodation_2017 <- read_excel(file.path(raw_path, "Accommodation_Capacity", "Capacità comunale 2013-2024.xlsx"),
sheet = "2017", col_names = FALSE, n_max = 15)
accommodation_2017
#####
# STEP 24D - Read 2017 Sheet with Headers
#####
accommodation_2017 <- read_excel(file.path(raw_path, "Accommodation_Capacity", "Capacità comunale 2013-2024.xlsx"),
sheet = "2017", skip = 4)
names(accommodation_2017)
head(accommodation_2017)
#####
# STEP 24E - Verify Total Beds Column
#####
accommodation_2017[, c("CODCOM...7", "LETTI / BEDS...9", "LETTI / BEDS...13", "LETTI / BEDS...17", "LETTI /
BEDS...21", "LETTI / BEDS...25", "LETTI / BEDS...29", "LETTI / BEDS...33", "LETTI / BEDS...37")] > head(10)
#####
# STEP 24F - Inspect All Bed Columns for One Municipality
#####
accommodation_2017[accommodation_2017$CODCOM...7 == "001003",]
#####
# STEP 24F - First Municipality with Beds
#####
accommodation_2017[accommodation_2017$`LETTI / BEDS...37` > 0,][1,]
#####
# STEP 24G - Display All Bed Columns
#####
print(accommodation_2017[accommodation_2017$CODCOM...7 == "001005", c("CODCOM...7", grep("LETTI / BEDS",
names(accommodation_2017), value = TRUE))], width = Inf)
#####
# STEP 25 - Create Accommodation Capacity Panel (2017-2024)
#####
years <- 2017:2024
accommodation_list <- list()
for (yr in years) {
  temp <- read_excel(
    file.path(raw_path, "Accommodation_Capacity", "Capacità comunale 2013-2024.xlsx"),
    sheet = as.character(yr),
    skip = 4
  )
}
```

```

)
temp <- as.data.table(temp)
temp <- temp[, .(Municipality_ID = `CODCOM...7`, Beds = `LETTI / BEDS...55`)]
temp[, Year := yr]
accommodation_list[[as.character(yr)]] <- temp
}
#####
# STEP 25A - Combine All Years
#####
accommodation_panel <- rbindlist(accommodation_list)
#####
# STEP 25B - Quality Checks
#####
dim(accommodation_panel)
head(accommodation_panel)
summary(accommodation_panel$Beds)
sum(is.na(accommodation_panel$Beds))
sum(duplicated(accommodation_panel[, .(Municipality_ID, Year)]))
#####
# STEP 25C - Inspect Duplicate Municipality-Year Records
#####
duplicate_rows <- accommodation_panel[duplicated(accommodation_panel, by = c("Municipality_ID", "Year")) |
duplicated(accommodation_panel, by = c("Municipality_ID", "Year"), fromLast = TRUE)]
duplicate_rows[order(Municipality_ID, Year)]
#####
# STEP 25D - Remove Summary Rows
#####
accommodation_panel <- accommodation_panel[!is.na(Municipality_ID)]
#####
# STEP 25E - Quality Checks
#####
dim(accommodation_panel)
sum(is.na(accommodation_panel$Municipality_ID))
sum(is.na(accommodation_panel$Beds))
sum(duplicated(accommodation_panel[, .(Municipality_ID, Year)]))
#####
# STEP 26 - Merge Accommodation Capacity
#####
final_panel <- merge(
  final_panel,
  accommodation_panel,
  by = c("Municipality_ID", "Year"),
  all.x = TRUE
)
#####
# STEP 26B - Quality Checks
#####
summary(final_panel$Beds)
summary(final_panel$TouristTaxPerBed)
sum(is.na(final_panel$Beds))
sum(is.na(final_panel$TouristTaxPerBed))
sum(duplicated(final_panel[, .(Municipality_ID, Year)]))
#####
# STEP 26A - Create Tourist Tax Revenue per Bed
#####
final_panel[, TouristTaxPerBed := fifelse(Beds > 0, TouristTaxRevenue / Beds, 0)]
#####
# STEP 26B - Quality Checks
#####
summary(final_panel$TouristTaxPerBed)
sum(is.na(final_panel$TouristTaxPerBed))
#####
# STEP 27 - Save Final Analysis Dataset
#####
save(final_panel, file = file.path(final_path, "Master_Analysis_Panel_Final_2017_2024.RData"))

```

```
#####
# STEP 27A - Export CSV
#####
fwrite(final_panel, file.path(final_path, "Master_Analysis_Panel_Final_2017_2024.csv"))
#####
# STEP 27B - Verify Output
#####
list.files(final_path)
#####
# STEP 27C - Final Dataset Summary
#####
dim(final_panel)
names(final_panel)
summary(final_panel)
```

A.1.2 Spatial Classification: Core, Neighbor, Second Ring and Other

```
#####
# A.1.2 Spatial Classification: Core, Neighbor, Second Ring and Other
#####
#####
# STEP 0: Set Working Directory
#####
setwd("D:/Edu/Masters/masters thesis/Master_Thesis_V3")
#####
# STEP 1: Load Required Packages
#####
library(dplyr)
library(data.table)
#####
# STEP 2: Load Final Analysis Dataset
#####
load("01_Data/Final_Data/Master_Analysis_Panel_Balanced_2017_2024.RData")
#####
# STEP 3: Check Loaded Dataset
#####
ls()
#####
# STEP 4: Create Yearly Intensity-Based Core Indicators
#####
balanced_panel <- balanced_panel %>% group_by(Year) %>% mutate(
  Core_Intensity_5_Year = ifelse(NightsPerResident >= quantile(NightsPerResident, probs = 0.95, na.rm = TRUE), 1, 0),
  Core_Intensity_10_Year = ifelse(NightsPerResident >= quantile(NightsPerResident, probs = 0.90, na.rm = TRUE), 1, 0)
) %>% ungroup()
#####
# STEP 5: Create Final Intensity-Based Core Variables
#####
intensity_summary <- balanced_panel %>% group_by(Municipality_ID) %>% summarise(
  Years_Core_5 = sum(Core_Intensity_5_Year), Years_Core_10 = sum(Core_Intensity_10_Year)
)
balanced_panel <- balanced_panel %>% left_join(intensity_summary, by = "Municipality_ID") %>% mutate(
  Core_Intensity_5 = ifelse(Years_Core_5 == 8, 1, 0),
  Core_Intensity_10 = ifelse(Years_Core_10 == 8, 1, 0)
)
#####
# STEP 6: Check Intensity-Based Core Definitions
#####
balanced_panel %>% distinct(Municipality_ID, Core_Intensity_5) %>% count(Core_Intensity_5)
balanced_panel %>% distinct(Municipality_ID, Core_Intensity_10) %>% count(Core_Intensity_10)
#####
# STEP 7: Create Yearly Volume-Based Core Indicators
#####
balanced_panel <- balanced_panel %>% group_by(Year) %>% mutate(
  Core_Volume_5_Year = ifelse(Nights >= quantile(Nights, probs = 0.95, na.rm = TRUE), 1, 0),
  Core_Volume_10_Year = ifelse(Nights >= quantile(Nights, probs = 0.90, na.rm = TRUE), 1, 0)
```

```

) %>% ungroup()
#####
# STEP 8: Create Final Volume-Based Core Variables
#####
volume_summary <- balanced_panel %>% group_by(Municipality_ID) %>% summarise(
  Years_Core_5 = sum(Core_Volume_5_Year), Years_Core_10 = sum(Core_Volume_10_Year)
)
balanced_panel <- balanced_panel %>% left_join(volume_summary, by = "Municipality_ID", suffix = c("", "_Volume"))
%>% mutate(
  Core_Volume_5 = ifelse(Years_Core_5_Volume == 8, 1, 0),
  Core_Volume_10 = ifelse(Years_Core_10_Volume == 8, 1, 0)
)
#####
# STEP 9: Check Volume-Based Core Definitions
#####
balanced_panel %>% distinct(Municipality_ID, Core_Volume_5) %>% count(Core_Volume_5)
balanced_panel %>% distinct(Municipality_ID, Core_Volume_10) %>% count(Core_Volume_10)
data.frame(Number = seq_along(names(balanced_panel)), Variable = names(balanced_panel))
data.frame(Number = seq_along(names(balanced_panel)), Variable = names(balanced_panel))
#####
# STEP 10: Create Union-Based Core Definitions
#####
balanced_panel <- balanced_panel %>% mutate(
  Core_Union_5 = ifelse(Core_Intensity_5 == 1 | Core_Volume_5 == 1, 1, 0),
  Core_Union_10 = ifelse(Core_Intensity_10 == 1 | Core_Volume_10 == 1, 1, 0)
)
#####
# STEP 11: Check Union-Based Core Definitions
#####
balanced_panel %>% distinct(Municipality_ID, Core_Union_5) %>% count(Core_Union_5)
balanced_panel %>% distinct(Municipality_ID, Core_Union_10) %>% count(Core_Union_10)
#####
# STEP 12: Compare Core Municipality Definitions
#####
table(Intensity = balanced_panel %>% distinct(Municipality_ID, Core_Intensity_5) %>% pull(Core_Intensity_5), Volume =
balanced_panel %>% distinct(Municipality_ID, Core_Volume_5) %>% pull(Core_Volume_5))
table(Intensity = balanced_panel %>% distinct(Municipality_ID, Core_Intensity_10) %>% pull(Core_Intensity_10),
Volume = balanced_panel %>% distinct(Municipality_ID, Core_Volume_10) %>% pull(Core_Volume_10))
#####
# STEP 13: Select Final Core Municipality Definition
#####
balanced_panel <- balanced_panel %>% mutate(Core_Final = Core_Union_5)
#####
# STEP 14: Compare Old and New Core Definitions
#####
table(Old_Core = balanced_panel %>% distinct(Municipality_ID, CoreDestination) %>% pull(CoreDestination), New_Core =
balanced_panel %>% distinct(Municipality_ID, Core_Union_5) %>% pull(Core_Union_5))
#####
# STEP 15: Load Municipality Shapefile
#####
library(sf)
municipality_sf <- st_read("01_Data/Raw_Data/Municipality_Registry/Com01012024/Com01012024_WGS84.shp", quiet =
TRUE)
names(municipality_sf)
#####
# STEP 16: Verify Municipality ID Match
#####
class(balanced_panel$Municipality_ID)
class(municipality_sf$PRO_COM_T)
sum(unique(balanced_panel$Municipality_ID) %in% municipality_sf$PRO_COM_T)
length(unique(balanced_panel$Municipality_ID))
nrow(municipality_sf)
#####
# STEP 17.1: Prepare Spatial Dataset
#####

```

```

core_data <- balanced_panel %>% distinct(Municipality_ID, Core_Final)
municipality_sf <- municipality_sf %>% left_join(core_data, by = c("PRO_COM_T" = "Municipality_ID"))
municipality_sf$Core_Final[is.na(municipality_sf$Core_Final)] <- 0
#####
# STEP 17.2: Identify Neighboring Municipalities
#####
touch_list <- st_touches(municipality_sf)
municipality_sf$NeighborCore_New <- 0
core_index <- which(municipality_sf$Core_Final == 1)
for(i in core_index){
  neighbors <- touch_list[[i]]
  neighbors <- neighbors[municipality_sf$Core_Final[neighbors] == 0]
  municipality_sf$NeighborCore_New[neighbors] <- 1
}
#####
# STEP 17.3: Remove Core Municipalities
#####
municipality_sf$NeighborCore_New[municipality_sf$Core_Final == 1] <- 0
#####
# STEP 17.4: Merge Back to Panel Dataset
#####
neighbor_data <- municipality_sf %>% st_drop_geometry() %>% select(PRO_COM_T, NeighborCore_New)
balanced_panel <- balanced_panel %>% left_join(neighbor_data, by = c("Municipality_ID" = "PRO_COM_T"))
#####
# STEP 17.5: Check Neighbor Classification
#####
balanced_panel %>% distinct(Municipality_ID, NeighborCore_New) %>% count(NeighborCore_New)
#####
# STEP 17.6: Validate Core and Neighbor Groups
#####
table(Core = municipality_sf$Core_Final, Neighbor = municipality_sf$NeighborCore_New)
#####
# STEP 18.1: Create 25 km Buffer Around Core Municipalities
#####
municipality_sf_utm <- st_transform(municipality_sf, 32632)
core_sf <- municipality_sf_utm %>% filter(Core_Final == 1)
core_buffer_25km <- st_buffer(core_sf, dist = 25000)
#####
# STEP 18.2: Merge Buffers
#####
core_buffer_25km <- st_union(core_buffer_25km)
#####
# STEP 18.3: Identify Municipalities Within 25 km
#####
inside_buffer <- st_intersects(municipality_sf_utm, core_buffer_25km, sparse = FALSE)
inside_buffer <- as.vector(inside_buffer)
municipality_sf_utm$SecondRing_25km <- as.integer(inside_buffer)
#####
# STEP 18.4: Remove Core and Neighbor Municipalities
#####
municipality_sf_utm$SecondRing_25km[municipality_sf_utm$Core_Final == 1] <- 0
municipality_sf_utm$SecondRing_25km[municipality_sf_utm$NeighborCore_New == 1] <- 0
#####
# STEP 18.5: Merge Back to Panel
#####
balanced_panel <- balanced_panel %>% select(-any_of("SecondRing_25km"))
second_ring_data <- municipality_sf_utm %>% st_drop_geometry() %>% select(PRO_COM_T, SecondRing_25km)
balanced_panel <- balanced_panel %>% left_join(second_ring_data, by = c("Municipality_ID" = "PRO_COM_T"))
#####
# STEP 18.6: Check Second-Ring Classification
#####
balanced_panel %>% distinct(Municipality_ID, SecondRing_25km) %>% count(SecondRing_25km)
#####
# STEP 19: Create Final Spatial Groups
#####

```

```

balanced_panel <- balanced_panel %>% mutate(
  Spatial_Group = case_when(
    Core_Final == 1 ~ "Core",
    NeighborCore_New == 1 ~ "Neighbor",
    SecondRing_25km == 1 ~ "Second Ring",
    TRUE ~ "Other"
  )
)
balanced_panel$Spatial_Group <- factor(balanced_panel$Spatial_Group, levels = c("Other", "Second Ring", "Neighbor",
"Core"))
#####
# STEP 19.1: Check Spatial Groups
#####
balanced_panel %>% distinct(Municipality_ID, Spatial_Group) %>% count(Spatial_Group)
#####
# STEP 20: Save Updated Analysis Dataset
#####
save(balanced_panel, file = "01_Data/Final_Data/Master_Analysis_Panel_Final_Spatial_2017_2024.RData")
save(balanced_panel, file = "01_Data/Final_Data/Master_Analysis_Panel_Final_Spatial_2017_2024.csv")

```

A.1.3 Extended Tourism Panel / Tourism Variable Construction

```

#####
# A.1.3 Extended Tourism Panel / Tourism Variable Construction
#####
# STEP 1 - Load Packages
#####
library(readxl)
library(data.table)
library(dplyr)
#####
# STEP 2 - Set Working Directory
#####
setwd("D:/Edu/Masters/masters thesis/Master_Thesis_V3")
#####
# STEP 3 - Define Paths
#####
raw_path <- "01_Data/Raw_Data/Tourism"
clean_path <- "01_Data/Clean_Data"
#####
# STEP 4 - Analysis Years
#####
years <- 2017:2024
#####
# STEP 5 - Empty List
#####
tourism_list <- list()
#####
# STEP 6 - Import Tourism Data
#####
for(i in seq_along(years)){
  year <- years[i]
  temp <- read_excel(file.path(raw_path, "2. Dati comunali 2014-2024.xlsx"), sheet = as.character(year), skip = 5)
  temp <- as.data.table(temp)
  temp[, Year := year]
  tourism_list[[i]] <- temp
}
#####
# STEP 6A - Combine Years
#####
tourism_extended <- rbindlist(tourism_list, fill = TRUE)
#####
# STEP 6B - Quality Checks
#####
dim(tourism_extended)

```

```

head(tourism_extended)
names(tourism_extended)
#####
# STEP 7 - Rename Tourism Variables
#####
setnames(tourism_extended, old = names(tourism_extended), new = c(
  "Region_Code", "Region", "Province_Code", "Province", "Municipality", "Municipality_ID", "Flag",
  "Resident_Arrivals", "Foreign_Arrivals", "Total_Arrivals",
  "Resident_Hotel_Arrivals", "Foreign_Hotel_Arrivals", "Total_Hotel_Arrivals",
  "Resident_ExtraHotel_Arrivals", "Foreign_ExtraHotel_Arrivals", "Total_ExtraHotel_Arrivals",
  "Blank", "Resident_Nights", "Foreign_Nights", "Total_Nights",
  "Resident_Hotel_Nights", "Foreign_Hotel_Nights", "Total_Hotel_Nights",
  "Resident_ExtraHotel_Nights", "Foreign_ExtraHotel_Nights", "Total_ExtraHotel_Nights", "Year"
))
#####
# STEP 7A - Verify Column Names
#####
names(tourism_extended)
head(tourism_extended)
#####
# STEP 8 - Replace Suppressed Values
#####
tourism_extended[tourism_extended == "(*)"] <- NA
#####
# STEP 8A - Convert Variables to Numeric
#####
numeric_cols <- c(
  "Resident_Arrivals", "Foreign_Arrivals", "Total_Arrivals",
  "Resident_Hotel_Arrivals", "Foreign_Hotel_Arrivals", "Total_Hotel_Arrivals",
  "Resident_ExtraHotel_Arrivals", "Foreign_ExtraHotel_Arrivals", "Total_ExtraHotel_Arrivals",
  "Resident_Nights", "Foreign_Nights", "Total_Nights",
  "Resident_Hotel_Nights", "Foreign_Hotel_Nights", "Total_Hotel_Nights",
  "Resident_ExtraHotel_Nights", "Foreign_ExtraHotel_Nights", "Total_ExtraHotel_Nights"
)
tourism_extended[, (numeric_cols) := lapply(.SD, as.numeric), .SDcols = numeric_cols]
#####
# STEP 8B - Quality Checks
#####
str(tourism_extended)
summary(tourism_extended[, ..numeric_cols])
sapply(tourism_extended[, ..numeric_cols], function(x) sum(is.na(x)))
#####
# STEP 9 - Foreign Tourist Share
#####
tourism_extended[, Foreign_Tourist_Share := fifelse(Total_Nights > 0, Foreign_Nights / Total_Nights, NA_real_)]
#####
# STEP 10 - Domestic Tourist Share
#####
tourism_extended[, Domestic_Tourist_Share := fifelse(Total_Nights > 0, Resident_Nights / Total_Nights, NA_real_)]
#####
# STEP 11 - Extra-Hotel Share
#####
tourism_extended[, ExtraHotel_Share := fifelse(Total_Nights > 0, Total_ExtraHotel_Nights / Total_Nights, NA_real_)]
#####
# STEP 12 - Hotel Share
#####
tourism_extended[, Hotel_Share := fifelse(Total_Nights > 0, Total_Hotel_Nights / Total_Nights, NA_real_)]
#####
# STEP 12A - Quality Checks
#####
summary(tourism_extended[, .(Foreign_Tourist_Share, Domestic_Tourist_Share, ExtraHotel_Share, Hotel_Share)])
sapply(tourism_extended[, .(Foreign_Tourist_Share, Domestic_Tourist_Share, ExtraHotel_Share, Hotel_Share)], function(x)
sum(is.na(x)))
#####
# STEP 13 - Save Extended Tourism Panel

```

```
#####
save(tourism_extended, file = file.path(clean_path, "Tourism_Panel_Extended_2017_2024.RData"))
fwrite(tourism_extended, file = file.path(clean_path, "Tourism_Panel_Extended_2017_2024.csv"))
#####
# STEP 13A - Verify Saved Files
#####
list.files(clean_path, pattern = "Tourism_Panel")
```

A.1.4 Chapter 2 - Final Balanced Panel Construction

```
#####
# A.1.4 Chapter 2 - Final Balanced Panel Construction
#####
# MASTER THESIS V3
# STEP 1 - Clean Workspace & Load Libraries
#####
# Clean workspace
rm(list = ls())
# Clear console
# Load libraries
library(data.table)
#####
# STEP 2 - Set Project Paths
#####
final_path <- "D:/Masters/masters thesis/Master_Thesis_V3/01_Data/Final_Data"
list.files("D:/Masters/masters thesis/Master_Thesis_V3/01_Data/Final_Data")
list.files("D:/Masters/masters thesis/Master_Thesis_V3")
#####
# STEP 3 - Load Final Unbalanced Panel
#####
load(file.path(final_path, "Master_Analysis_Panel_Final_2017_2024"))
setDT(final_panel)
#####
# STEP 4 - Verify Panel Before Balancing
#####
print(nrow(final_panel))
print(uniqueN(final_panel$Municipality_ID))
print(sort(unique(final_panel$Year)))
print(final_panel[, .N, by = Year][order(Year)])
#####
# STEP 5 - Count Years per Municipality
#####
municipality_count <- final_panel[, (Years_Observed = uniqueN(Year)), by = Municipality_ID]
#####
# STEP 6 - Keep Municipalities Present in All Years
#####
balanced_ids <- municipality_count[Years_Observed == 8, Municipality_ID]
#####
# STEP 7 - Create Balanced Panel
#####
balanced_panel <- final_panel[Municipality_ID %in% balanced_ids]
#####
# STEP 8 - Sort Dataset
#####
setorder(balanced_panel, Municipality_ID, Year)
#####
# STEP 9 - Quality Checks
#####
print(nrow(balanced_panel))
print(uniqueN(balanced_panel$Municipality_ID))
print(sort(unique(balanced_panel$Year)))
print(balanced_panel[, .N, by = Year][order(Year)])
print(sum(duplicated(balanced_panel[, (Municipality_ID, Year)])))
n_municipalities <- uniqueN(balanced_panel$Municipality_ID)
n_years <- uniqueN(balanced_panel$Year)
```

```

print(n_municipalities)
print(n_years)
print(n_municipalities * n_years)
#####
# STEP 10 - Save Balanced Panel
#####
save(balanced_panel,file = file.path(final_path,"Master_Analysis_Panel_Balanced_2017_2024.RData"))
fwrite(balanced_panel,file = file.path(final_path,"Master_Analysis_Panel_Balanced_2017_2024.csv"))
#####
# STEP 11 - Verify Saved Files
#####
list.files(final_path,pattern = "Balanced")

```

A.2 Chapters 3–4 — Empirical Analysis

A.2.1 ANOVA and Tukey HSD

```

#####
# ECONOMETRIC ANALYSIS — PROFESSOR REVISION
# Master Thesis
#####
#####
# STEP 0 - Clear Workspace
#####
rm(list = ls())
cat("\014")
#####
# STEP 1: Load Required Packages
#####
library(dplyr)
library(sf)
library(spdep)
library(lmtest)
library(sandwich)
#####
# STEP 2: Load Final Spatial Analysis Dataset
#####
load("D:/Edu/Masters/masters
thesis/Master_Thesis_V3/01_Data/Final_Data/Master_Analysis_Panel_Final_Spatial_2017_2024.RData")
#####
# STEP 3: Check Loaded Objects
#####
ls()
#####
# STEP 4: Check Panel Structure
#####
balanced_panel %>% summarise(Observations = n(), Municipalities = n_distinct(Municipality_ID), Years =
n_distinct(Year), Min_Year = min(Year, na.rm = TRUE), Max_Year = max(Year, na.rm = TRUE))
#####
# STEP 5: Verify Pre-Defined Spatial Groups
#####
spatial_group_check <- balanced_panel %>% distinct(Municipality_ID, Spatial_Group) %>% count(Spatial_Group) %>%
arrange(factor(Spatial_Group, levels = c("Core", "Neighbor", "Second Ring", "Other")))
spatial_group_check
#####
# STEP 6: Check Variables for ANOVA
#####
names(balanced_panel)
summary(balanced_panel %>% select(TouristTaxPerResident, TouristTaxPerNight, NightsPerResident))
#####
# STEP 7: Create Municipality-Level 2017–2024 Averages
#####
anova_data <- balanced_panel %>% group_by(Municipality_ID, Spatial_Group) %>%
summarise(Mean_TouristTaxPerResident = mean(TouristTaxPerResident, na.rm = TRUE), Mean_TouristTaxPerNight =
mean(TouristTaxPerNight, na.rm = TRUE), Mean_NightsPerResident = mean(NightsPerResident, na.rm = TRUE), .groups
= "drop")

```

```
#####
# STEP 8: Check ANOVA Dataset
#####
anova_data %>% summarise(Municipalities = n_distinct(Municipality_ID), Observations = n(), Groups =
n_distinct(Spatial_Group))
#####
# STEP 9: Check Municipalities per Spatial Group
#####
anova_data %>% count(Spatial_Group)
#####
# STEP 10: Check Missing Values in ANOVA Dataset
#####
anova_data %>% summarise(Missing_TaxPerResident = sum(is.na(Mean_TouristTaxPerResident)), Missing_TaxPerNight
= sum(is.na(Mean_TouristTaxPerNight)), Missing_NightsPerResident = sum(is.na(Mean_NightsPerResident)))
#####
# STEP 11: ANOVA — Tourist Tax Revenue per Resident
#####
anova_tax_resident <- aov(Mean_TouristTaxPerResident ~ Spatial_Group, data = anova_data)
summary(anova_tax_resident)
#####
# STEP 12: ANOVA — Tourist Tax Revenue per Night
#####
anova_tax_night <- aov(Mean_TouristTaxPerNight ~ Spatial_Group, data = anova_data)
summary(anova_tax_night)
#####
# STEP 13: ANOVA — Nights per Resident
#####
anova_nights_resident <- aov(Mean_NightsPerResident ~ Spatial_Group, data = anova_data)
summary(anova_nights_resident)
#####
# STEP 14: Tukey HSD — Tourist Tax Revenue per Resident
#####
tukey_tax_resident <- TukeyHSD(anova_tax_resident)
tukey_tax_resident
#####
# STEP 15: Tukey HSD — Tourist Tax Revenue per Night
#####
tukey_tax_night <- TukeyHSD(anova_tax_night)
tukey_tax_night
#####
# STEP 16: Tukey HSD — Nights per Resident
#####
tukey_nights_resident <- TukeyHSD(anova_nights_resident)
tukey_nights_resident
#####
# STEP 17: Create Clean Tukey HSD Results
#####
tukey_tax_resident_table <- as.data.frame(tukey_tax_resident$Spatial_Group) %>%
tibble::rownames_to_column("Comparison")
tukey_tax_night_table <- as.data.frame(tukey_tax_night$Spatial_Group) %>% tibble::rownames_to_column("Comparison")
tukey_nights_resident_table <- as.data.frame(tukey_nights_resident$Spatial_Group) %>%
tibble::rownames_to_column("Comparison")
#####
# STEP 18: Display Tukey Tables
#####
tukey_tax_resident_table
tukey_tax_night_table
tukey_nights_resident_table
#####
# SAVE ANOVA & TUKEY RESULTS
#####
output_path <- "D:/Edu/Masters/masters thesis/Master_Thesis_V3/03_Output/Tables"
# ANOVA results
write.csv(as.data.frame(summary(anova_tax_resident)[[1]]), file.path(output_path, "ANOVA_TouristTaxPerResident.csv"))
write.csv(as.data.frame(summary(anova_tax_night)[[1]]), file.path(output_path, "ANOVA_TouristTaxPerNight.csv"))
```

```

write.csv(as.data.frame(summary(anova_nights_resident)[[1]]), file.path(output_path, "ANOVA_NightsPerResident.csv"))
# Tukey results
write.csv(tukey_tax_resident_table, file.path(output_path, "Tukey_TouristTaxPerResident.csv"), row.names = FALSE)
write.csv(tukey_tax_night_table, file.path(output_path, "Tukey_TouristTaxPerNight.csv"), row.names = FALSE)
write.csv(tukey_nights_resident_table, file.path(output_path, "Tukey_NightsPerResident.csv"), row.names = FALSE)
#####
# Confirm
#####
list.files(output_path, pattern = "ANOVA|Tukey", full.names = FALSE)

```

A.2.2 Global Moran's I

```

#####
# 04_02_Global_Morans_I_Final.R
# Final Global Moran's I Analysis
#
# Purpose:
# - Measure global spatial autocorrelation in the main
#   tourism and tourist-tax variables.
# - Create the spatial weights matrix using Queen
#   contiguity with a nearest-neighbour fallback.
# - Estimate Global Moran's I using municipality-level
#   averages for 2017–2024.
# - Save the final results used in the thesis.
# Spatial unit: Municipality
# Analysis period: 2017–2024
#####
# STEP 1: Load Required Packages
#####
library(dplyr)
library(sf)
library(spdep)
#####
# STEP 2: Load the Final Analysis Dataset
#####
load("D:/Edu/Masters/masters
thesis/Master_Thesis_V3/01_Data/Final_Data/Master_Analysis_Panel_Final_Spatial_2017_2024.RData")
#####
# STEP 3: Calculate Municipality-Level Averages (2017–2024)
#####
moran_data <- balanced_panel %>% group_by(Municipality_ID) %>% summarise(NightsPerResident =
mean(NightsPerResident, na.rm = TRUE), TouristTaxRevenue = mean(TouristTaxRevenue, na.rm = TRUE),
TouristTaxPerResident = mean(TouristTaxPerResident, na.rm = TRUE), TouristTaxPerNight = mean(TouristTaxPerNight,
na.rm = TRUE), .groups = "drop")
#####
# STEP 4: Load and Match Municipality Boundaries
#####
municipalities_sf <- st_read("D:/Edu/Masters/masters
thesis/Master_Thesis_V3/01_Data/Raw_Data/Municipality_Registry/Com01012024/Com01012024_WGS84.shp", quiet =
TRUE) %>% mutate(Municipality_ID = PRO_COM_T) %>% semi_join(moran_data, by = "Municipality_ID")
# Check that all municipalities in the analysis dataset
# have a matching spatial boundary.
stopifnot(nrow(municipalities_sf) == n_distinct(moran_data$Municipality_ID))
#####
# STEP 5: Create Queen Contiguity Neighbours
#####
neighbors <- poly2nb(municipalities_sf, queen = TRUE)
# Identify municipalities with no Queen neighbours.
zero_neighbour_ids <- which(card(neighbors) == 0)
#####
# STEP 6: Add Nearest-Neighbour Fallback
#####
# Calculate municipality centroids.
municipality_centroids <- st_centroid(municipalities_sf)
# Extract centroid coordinates.

```

```

centroid_coords <- st_coordinates(municipality_centroids)
# Find the nearest municipality for each municipality.
nearest_neighbour <- knn2nb(knearneigh(centroid_coords, k = 1))
# Replace the empty Queen-neighbour list for municipalities
# that have no Queen neighbours. for (i in zero_neighbour_ids) {neighbors[[i]] <- nearest_neighbour[[i]]}
# Confirm that every municipality now has at least
# one spatial neighbour.
stopifnot(all(card(neighbors) > 0))
#####
# STEP 7: Create Row-Standardized Spatial Weights
#####
spatial_weights <- nb2listw(neighbors, style = "W", zero.policy = TRUE)
#####
# STEP 8: Run Global Moran's I Tests
#####
moran_nights <- moran.test(moran_data$NightsPerResident, spatial_weights, zero.policy = TRUE, randomisation = TRUE,
alternative = "greater")
moran_tax <- moran.test(moran_data$TouristTaxRevenue, spatial_weights, zero.policy = TRUE, randomisation = TRUE,
alternative = "greater")
moran_tax_resident <- moran.test(moran_data$TouristTaxPerResident, spatial_weights, zero.policy = TRUE, randomisation
= TRUE, alternative = "greater")
moran_tax_night <- moran.test(moran_data$TouristTaxPerNight, spatial_weights, zero.policy = TRUE, randomisation =
TRUE, alternative = "greater")
#####
# STEP 8A: Verify Moran's I Test Configuration
#####
moran_nights
moran_tax
moran_tax_resident
moran_tax_night
#####
# STEP 9: Create the Moran's I Results Table
#####
moran_results <- data.frame(Variable = c("Nights per Resident", "Tourist Tax Revenue", "Tourist Tax per Resident",
"Tourist Tax per Night"), Moran_I = c(moran_nights$estimate[1], moran_tax$estimate[1], moran_tax_resident$estimate[1],
moran_tax_night$estimate[1]), P_Value = c(moran_nights$p.value, moran_tax$p.value, moran_tax_resident$p.value,
moran_tax_night$p.value))
#####
# STEP 10: Format and Display the Results
#####
moran_results <- moran_results %>% mutate(Moran_I = round(Moran_I, 3), P_Value = ifelse(P_Value < 0.001, "<0.001",
sprintf("%.3f", P_Value)))
moran_results
#####
# STEP 11: Save the Final Results
#####
output_path <- "D:/Edu/Masters/masters thesis/Master_Thesis_V3/03_Output/Tables"
dir.create(output_path, recursive = TRUE, showWarnings = FALSE)
write.csv(moran_results, file.path(output_path, "Table_Morans_I_Results.csv"), row.names = FALSE)
#####
# STEP 12: Verify Current Spatial Weights
#####
sum(card(neighbors))
mean(card(neighbors))
summary(card(neighbors))
sum(card(neighbors) == 0)
length(neighbors)
#####
# STEP 13: Verify Row-Standardization
#####
row_sums <- sapply(spatial_weights$weights, sum)
min(row_sums)
max(row_sums)
sum(abs(row_sums - 1) > 1e-10)
#####

```

```
# END OF SECTION 4.2
#####
length(zero_neighbour_ids))
```

A.2.3 Baseline Spatial-Group Regression

```
#####
# 04_03_Baseline_Spatial_Group_Regression_Final.R
# Final Baseline Spatial-Group Regression Analysis
#
# Purpose:
# - Examine whether tourist-tax outcomes differ
#   systematically across spatial groups.
# - Compare Core, Neighbor, Second Ring and Other municipalities.
# - Include year effects and municipal controls.
# - Use municipality-clustered standard errors.
# The analysis is associative, not causal.
# Analysis period:2017–2024
#####
# STEP 1: Clear Environment
#####
rm(list = ls())
gc()
#####
# STEP 2: Load Required Packages
#####
library(dplyr)
library(sandwich)
library(sf)
#####
# STEP 3: Set Project Path
#####
setwd("D:/Edu/Masters/masters thesis/Master_Thesis_V3")
#####
# STEP 4: Load Final Analysis Dataset
#####
load("01_Data/Final_Data/Master_Analysis_Panel_Final_Spatial_2017_2024.RData")
#####
# STEP 5: Check Spatial-Group Composition
#####
table(balanced_panel$Spatial_Group)
#####
# STEP 6: Set Core as Reference Group
#####
balanced_panel <- balanced_panel %>% mutate(Spatial_Group = relevel(factor(Spatial_Group), ref = "Core"))
#####
# STEP 7: Verify Reference Category
#####
levels(balanced_panel$Spatial_Group)
#####
# STEP 8: Check Final Model Variables
#####
summary(balanced_panel %>% select(TouristTaxRevenue, NightsPerResident, Population, Beds, Foreign_Tourist_Share,
Spatial_Group, Year, Municipality_ID))
#####
# STEP 9: Define the Final Baseline Specification
#####
# Core municipalities are the reference group.
# Year fixed effects account for common changes across municipalities over time.
# Population, Beds, Nights per Resident and Foreign Tourist Share are included as controls.
# ExtraHotel_Share is excluded because approximately 45% of observations are missing.
# The model is associative, not causal.
#####
#####
# STEP 9A: Load Municipality Registry for PRO_COM_T
```

```
#####
municipalities_sf <- st_read("01_Data/Raw_Data/Municipality_Registry/Com01012024/Com01012024_WGS84.shp", quiet
= TRUE)
#####
# STEP 9B: Verify PRO_COM_T
#####
names(municipalities_sf)[grep("PRO_COM_T", names(municipalities_sf))]
n_distinct(municipalities_sf$PRO_COM_T)
#####
# STEP 9C: Create PRO_COM_T Municipality Lookup
#####
pro_com_lookup <- municipalities_sf %>% st_drop_geometry() %>% select(PRO_COM_T) %>% distinct()
#####
# STEP 9D: Inspect PRO_COM_T Structure
#####
head(pro_com_lookup)
str(pro_com_lookup$PRO_COM_T)
n_distinct(pro_com_lookup$PRO_COM_T)
#####
# STEP 9E: Inspect Existing Panel Municipality_ID
#####
head(balanced_panel$Municipality_ID)
str(balanced_panel$Municipality_ID)
n_distinct(balanced_panel$Municipality_ID)
#####
# STEP 9F: Verify Municipality_ID Matches PRO_COM_T
#####
panel_ids <- balanced_panel %>% distinct(Municipality_ID)
sum(panel_ids$Municipality_ID %in% pro_com_lookup$PRO_COM_T)
nrow(panel_ids)
setdiff(panel_ids$Municipality_ID, pro_com_lookup$PRO_COM_T)
#####
# STEP 9G: Create Municipality_ID -> PRO_COM_T Lookup
#####
pro_com_lookup <- municipalities_sf %>% st_drop_geometry() %>% select(PRO_COM_T) %>% distinct() %>%
mutate(Municipality_ID = PRO_COM_T) %>% select(Municipality_ID, PRO_COM_T)
#####
# STEP 9H: Verify Lookup Uniqueness
#####
pro_com_lookup %>% count(Municipality_ID) %>% filter(n > 1)
#####
# STEP 9I: Add PRO_COM_T to balanced_panel
#####
balanced_panel <- balanced_panel %>% left_join(pro_com_lookup, by = "Municipality_ID")
#####
# STEP 9J: Verify PRO_COM_T in Final Panel
#####
n_distinct(balanced_panel$Municipality_ID)
n_distinct(balanced_panel$PRO_COM_T)
sum(is.na(balanced_panel$PRO_COM_T))
head(balanced_panel %>% select(Municipality_ID, PRO_COM_T))
#####
# STEP 10: Estimate Tourist Tax Revenue Model
#####
model_tax_revenue <- lm(TouristTaxRevenue ~ Spatial_Group + factor(Year) + Population + Beds + NightsPerResident +
Foreign_Tourist_Share, data = balanced_panel)
#####
# STEP 11: Calculate Municipality-Clustered
# Standard Errors
#####
library(lmtest)
V_cluster <- vcovCL(model_tax_revenue, cluster = ~ PRO_COM_T, type = "HC1")
coefTest(model_tax_revenue, vcov = V_cluster)
#####
# STEP 12: Calculate Clustered Test Statistics
```

```

# and p-values
#####
create_clustered_table <- function(model, clustered_vcov) {
  estimates <- coef(model)
  clustered_se <- sqrt(diag(clustered_vcov))
  clustered_t <- estimates / clustered_se
  clustered_p <- 2 * pt(abs(clustered_t), df = n_distinct(balanced_panel$Municipality_ID) - 1, lower.tail = FALSE)
  data.frame(Variable = names(estimates), Estimate = estimates, Clustered_SE = clustered_se, Clustered_t = clustered_t,
P_Value = clustered_p, row.names = NULL)
}
#####
# STEP 13: Create Final Regression Results Table
#####
results_tax_revenue <- create_clustered_table(model_tax_revenue, V_cluster)
#####
# STEP 14: Add Model Statistics
#####
add_model_statistics <- function(results_table, model) {
  model_statistics <- data.frame(Variable = c("Observations", "R-squared", "Adjusted R-squared"), Estimate =
c(nobs(model), summary(model)$r.squared, summary(model)$adj.r.squared), Clustered_SE = NA, Clustered_t = NA,
P_Value = NA)
  bind_rows(results_table, model_statistics)
}
results_tax_revenue <- add_model_statistics(results_tax_revenue, model_tax_revenue)
#####
# STEP 17: Clean Variable Labels
#####
clean_labels <- function(results_table) {
  results_table %>%
  mutate(
    Variable = case_when(
      Variable == "(Intercept)" ~ "Constant",
      Variable == "Spatial_GroupOther" ~ "Other",
      Variable == "Spatial_GroupSecond Ring" ~ "Second Ring",
      Variable == "Spatial_GroupNeighbor" ~ "Neighbor",
      Variable == "factor(Year)2018" ~ "2018",
      Variable == "factor(Year)2019" ~ "2019",
      Variable == "factor(Year)2020" ~ "2020",
      Variable == "factor(Year)2021" ~ "2021",
      Variable == "factor(Year)2022" ~ "2022",
      Variable == "factor(Year)2023" ~ "2023",
      Variable == "factor(Year)2024" ~ "2024",
      Variable == "Population" ~ "Population",
      Variable == "Beds" ~ "Beds",
      Variable == "NightsPerResident" ~ "Nights per Resident",
      Variable == "Foreign_Tourist_Share" ~ "Foreign Tourist Share",
      TRUE ~ Variable
    )
  ) %>%
  mutate(
    Estimate = ifelse(Variable %in% c("Observations", "R-squared", "Adjusted R-squared"), round(Estimate, 4),
round(Estimate, 3)),
    Clustered_SE = round(Clustered_SE, 3),
    Clustered_t = round(Clustered_t, 3),
    P_Value = ifelse(is.na(P_Value), NA, ifelse(P_Value < 0.001, "<0.001", sprintf("%.3f", P_Value)))
  )
}
results_tax_revenue <- clean_labels(results_tax_revenue)
#####
# STEP 18: Display Final Results
#####
print(results_tax_revenue)
#####
# STEP 19: Save Final Results
#####

```

```

output_path <- "03_Output/Tables"
dir.create(output_path, recursive = TRUE, showWarnings = FALSE)
write.csv(results_tax_revenue, file.path(output_path, "Table_Baseline_Spatial_Group_TouristTaxRevenue_Final.csv"),
row.names = FALSE)
#####
# STEP 20: Run Final Verification Checks
#####
# Number of municipalities
n_distinct(balanced_panel$Municipality_ID)
# Number of observations in the model
nobs(model_tax_revenue)
# Model fit
summary(model_tax_revenue)$r.squared
summary(model_tax_revenue)$adj.r.squared
#####
# STEP 21: End of Section 3.3
#####

```

A.2.4 Core-Cluster Regression

A.2.4.1 Core-Cluster Creation and Assignment

```

#####
# 04_00_Core_Cluster_Assignment_Final.R
# FINAL CORE-CLUSTER CONSTRUCTION
# Purpose:
# 1. Identify the 162 Core municipalities
# 2. Construct 81 Core clusters using Queen contiguity
# 3. Assign all 481 Neighbor municipalities to Core clusters
# 4. Assign Second Ring municipalities using the SAME 25-km geographical rule used in the original
# Second Ring classification
# 5. Use centroid distance ONLY as a tie-breaker when multiple Core clusters qualify
# 6. Create and save the final Core-cluster lookup
#
# IMPORTANT:
# Second Ring eligibility is NOT determined by centroid-to-centroid distance.
#####
# STEP 1: Load Packages
#####
library(dplyr)
library(sf)
library(data.table)
#####
# STEP 2: Set Project Directory
#####
setwd("D:/Edu/Masters/masters thesis/Master_Thesis_V3")
#####
# STEP 3: Load Final Balanced Panel
#####
load("01_Data/Final_Data/Master_Analysis_Panel_Final_Spatial_2017_2024.RData")
#####
# STEP 4: Basic Panel Verification
#####
balanced_panel %>%
  distinct(Municipality_ID, Spatial_Group) %>%
  count(Spatial_Group)
#####
# STEP 5: Create Municipality-Level Spatial Lookup
#####
spatial_lookup <- balanced_panel %>%
  select(Municipality_ID, Municipality, Spatial_Group) %>%
  distinct()
#####
# STEP 6: Validate Final Spatial Classification

```

```
#####
#-----
# STEP 6A: Check panel dimensions
#-----
dim(balanced_panel)
#-----
# STEP 6B: Check number of unique municipalities
#-----
n_distinct(balanced_panel$Municipality_ID)
#-----
# STEP 6C: Check that each municipality has only ONE
# spatial-group assignment
#-----
balanced_panel %>%
  select(Municipality_ID, Spatial_Group) %>%
  distinct() %>%
  count(Municipality_ID) %>%
  filter(n > 1)
#-----
# STEP 6D: Check spatial-group counts
#-----
balanced_panel %>%
  select(Municipality_ID, Spatial_Group) %>%
  distinct() %>%
  count(Spatial_Group)
#-----
# STEP 6E: Check whether spatial-group assignment is
# constant across all years
#-----
balanced_panel %>%
  select(Municipality_ID, Year, Spatial_Group) %>%
  distinct() %>%
  group_by(Municipality_ID) %>%
  summarise(
    n_groups = n_distinct(Spatial_Group),
    groups = paste(unique(Spatial_Group), collapse = " | "),
    n_years = n_distinct(Year),
    .groups = "drop"
  ) %>%
  count(n_groups, groups)
#####
# STEP 7: Extract and Verify Core Municipalities
#####
# STEP 7A: Extract Core municipalities from the final
# spatial classification
#-----
core_ids <- balanced_panel %>%
  filter(Spatial_Group == "Core") %>%
  distinct(Municipality_ID, .keep_all = TRUE) %>%
  select(Municipality_ID, Municipality)
#-----
# STEP 7B: Verify number of Core municipalities
#-----
nrow(core_ids)
#-----
# STEP 7C: Check for duplicate Core municipality IDs
#-----
core_ids %>%
  count(Municipality_ID) %>%
  filter(n > 1)
#-----
# STEP 7D: Display Core municipality count
#-----
core_ids %>%
  summarise(
```

```

Core_Municipalities = n_distinct(Municipality_ID)
)
#####
# STEP 8: Construct Core Clusters Using Queen Contiguity
#####
#-----
# STEP 8A: Prepare municipality shapefile
#-----
municipalities_sf <- st_read(
  "01_Data/Raw_Data/Municipality_Registry/Com01012024/Com01012024_WGS84.shp", quiet = TRUE)
#-----
# STEP 8C: Create Municipality_ID in the shapefile
#-----
municipalities_sf <- municipalities_sf %>%
  mutate(Municipality_ID = PRO_COM_T )
#####
# END OF STEP 8 PREPARATION
#####
#-----
# STEP 8C: Check coordinate reference system
#-----
st_crs(municipalities_sf)
#-----
# STEP 8D: Transform to projected CRS for spatial analysis
#-----
municipalities_sf <- municipalities_sf %>%
  st_transform(32632)
#-----
# STEP 8E: Keep only the 162 Core municipalities
#-----
core_sf <- municipalities_sf %>%
  inner_join(
    core_ids,
    by = "Municipality_ID")
#-----
# STEP 8F: Verify Core spatial object
#-----
nrow(core_sf)
n_distinct(core_sf$Municipality_ID)
#-----
# STEP 8G: Construct Queen-contiguity neighbours
#-----
core_nb <- poly2nb( core_sf, queen = TRUE)
#-----
# STEP 8H: Check Core neighbour structure
#-----
summary(core_nb)
#-----
# STEP 8I: Identify connected Core components
#-----
core_components <- n.comp.nb(core_nb)
#-----
# STEP 8J: Create Core Cluster ID
#-----
core_sf$Core_Cluster_ID <- core_components$comp.id
#-----
# STEP 8K: Count Core clusters
#-----
n_core_clusters <- n_distinct(core_sf$Core_Cluster_ID)
n_core_clusters
#-----
# STEP 8L: Check Core municipalities per cluster
#-----
core_sf %>%
  st_drop_geometry() %>%

```

```

count(Core_Cluster_ID, name = "N_Core_Municipalities") %>%
  arrange(Core_Cluster_ID)
#-----
# STEP 8M: Final structural check
#-----
stopifnot(
  nrow(core_sf) == 162,
  n_distinct(core_sf$Municipality_ID) == 162,
  n_core_clusters == 81)
#####
# STEP 9: Validate Core-Cluster Structure
#####
#-----
# STEP 9A: Check every Core municipality has one cluster ID
#-----
core_sf %>%
  st_drop_geometry() %>%
  select(
    Municipality_ID,
    Core_Cluster_ID
  ) %>%
  count(Municipality_ID) %>%
  filter(n > 1)
#-----
# STEP 9B: Check for missing Core-cluster IDs
#-----
sum(is.na(core_sf$Core_Cluster_ID))
#-----
# STEP 9C: Verify number of Core clusters
#-----
n_distinct(core_sf$Core_Cluster_ID)
#-----
# STEP 9D: Verify total Core municipalities
#-----
n_distinct(core_sf$Municipality_ID)
#-----
# STEP 9E: Check cluster-size distribution
#-----
core_sf %>%
  st_drop_geometry() %>%
  count(Core_Cluster_ID, name = "N_Core_Municipalities") %>%
  summarise(
    Number_of_Clusters = n(),
    Total_Core_Municipalities = sum(N_Core_Municipalities),
    Minimum_Cluster_Size = min(N_Core_Municipalities),
    Maximum_Cluster_Size = max(N_Core_Municipalities)
  )
#-----
# STEP 9F: Final validation
#-----
stopifnot(
  n_distinct(core_sf$Municipality_ID) == 162,
  n_distinct(core_sf$Core_Cluster_ID) == 81,
  sum(is.na(core_sf$Core_Cluster_ID)) == 0)
#####
# STEP 10: Assign Neighbor Municipalities to Core Clusters
#####
#-----
# STEP 10A: Extract the 481 Neighbor municipalities
#-----
neighbor_ids <- balanced_panel %>%
  filter(Spatial_Group == "Neighbor") %>%
  distinct(Municipality_ID, .keep_all = TRUE) %>%
  select(Municipality_ID, Municipality)
#-----

```

```

# STEP 10B: Verify Neighbor count
#-----
nrow(neighbor_ids)
n_distinct(neighbor_ids$Municipality_ID)
#-----
# STEP 10C: Create spatial object for Neighbor municipalities
#-----
neighbor_sf <- municipalities_sf %>%
  inner_join(
    neighbor_ids,
    by = "Municipality_ID" )
#-----
# STEP 10D: Verify Neighbor spatial object
#-----
nrow(neighbor_sf)
n_distinct(neighbor_sf$Municipality_ID)
#-----
# STEP 10F: Identify Core municipalities that directly
# border each Neighbor municipality
#-----
neighbor_core_links <- lapply(
  seq_len(nrow(neighbor_sf)),
  function(i) {
    core_matches <- st_intersects(
      neighbor_sf[i, ],
      core_sf,
      sparse = TRUE
    )[[1]]
    if (length(core_matches) == 0) {
      return(NULL)
    }
    tibble(
      Neighbor_Municipality_ID = neighbor_sf$Municipality_ID[i],
      Core_Municipality_ID = core_sf$Municipality_ID[core_matches],
      Core_Cluster_ID = core_sf$Core_Cluster_ID[core_matches]
    )
  }
) %>%
  bind_rows()
#-----
# STEP 10G: Check whether every Neighbor has at least
# one qualifying Core municipality
#-----
neighbor_core_coverage <- neighbor_ids %>%
  left_join(
    neighbor_core_links %>%
      distinct(Neighbor_Municipality_ID),
    by = c("Municipality_ID" = "Neighbor_Municipality_ID")
  )
sum(is.na(neighbor_core_coverage$Core_Municipality_ID))
#-----
# STEP 10H: Count qualifying Core clusters per Neighbor
#-----
neighbor_cluster_candidates <- neighbor_core_links %>%
  group_by(Neighbor_Municipality_ID) %>%
  summarise(
    n_qualifying_core = n_distinct(Core_Municipality_ID),
    n_qualifying_clusters = n_distinct(Core_Cluster_ID),
    .groups = "drop"
  )
#-----
# STEP 10I: Check distribution of qualifying Core clusters
#-----
neighbor_cluster_candidates %>%
  count(

```

```

    n_qualifying_core,
    n_qualifying_clusters,
    sort = TRUE
  )
  neighbor_ids %>%
    anti_join(
      neighbor_core_links %>%
        distinct(Neighbor_Municipality_ID),
      by = c("Municipality_ID" = "Neighbor_Municipality_ID")
    )
#####
# STEP 10K: Assign Neighbor Municipalities to Core Clusters
#####
#-----
# STEP 10K-1: Calculate centroid distance for each
# Neighbor-Core candidate pair
#-----
neighbor_centroids <- st_centroid(neighbor_sf)
core_centroids <- st_centroid(core_sf)
#-----
# STEP 10K-2: Calculate distances only between each Neighbor
# and its qualifying Core municipalities
#-----
neighbor_core_distances <- neighbor_core_links %>%
  rowwise() %>%
  mutate(
    Distance_m = as.numeric(
      st_distance(
        neighbor_centroids[
          which(
            neighbor_centroids$Municipality_ID ==
              Neighbor_Municipality_ID
          ),
        ],
        core_centroids[
          which(
            core_centroids$Municipality_ID ==
              Core_Municipality_ID
          ),
        ]
      )
    )
  ) %>%
  ungroup()
#-----
# STEP 10K-3: Assign the nearest qualifying Core municipality
#-----
neighbor_cluster_assignment <- neighbor_core_distances %>%
  arrange(
    Neighbor_Municipality_ID,
    Distance_m,
    Core_Cluster_ID,
    Core_Municipality_ID
  ) %>%
  group_by(Neighbor_Municipality_ID) %>%
  slice(1) %>%
  ungroup() %>%
  select(
    Neighbor_Municipality_ID,
    Core_Municipality_ID,
    Core_Cluster_ID,
    Distance_m
  )
#-----
# STEP 10K-4: Verify one assignment per Neighbor

```

```

#-----
neighbor_cluster_assignment %>%
  count(Neighbor_Municipality_ID) %>%
  filter(n > 1)
#-----
# STEP 10K-5: Verify all 481 Neighbors were assigned
#-----
nrow(neighbor_cluster_assignment)
n_distinct(
  neighbor_cluster_assignment$Neighbor_Municipality_ID
)
#-----
# STEP 10K-6: Check for missing Core-cluster assignments
#-----
sum(
  is.na(
    neighbor_cluster_assignment$Core_Cluster_ID
  )
)
#####
# STEP 11: Assign Second Ring Municipalities
#####
#-----
# STEP 11A: Extract the 1,171 Second Ring municipalities
#-----
second_ring_ids <- balanced_panel %>%
  filter(Spatial_Group == "Second Ring") %>%
  distinct(Municipality_ID, .keep_all = TRUE) %>%
  select(
    Municipality_ID,
    Municipality
  )
#-----
# STEP 11B: Verify Second Ring count
#-----
nrow(second_ring_ids)
n_distinct(second_ring_ids$Municipality_ID)
#-----
# STEP 11C: Create spatial object for Second Ring
#-----
second_ring_sf <- municipalities_sf %>%
  inner_join(
    second_ring_ids,
    by = "Municipality_ID"
  )
#-----
# STEP 11D: Verify Second Ring spatial object
#-----
nrow(second_ring_sf)
n_distinct(second_ring_sf$Municipality_ID)
#####
# STEP 11E: Create 25-km buffers around each Core
#####
core_buffers_25km <- st_buffer(
  core_sf,
  dist = 25000
)
#-----
# STEP 11F: Identify Core municipalities satisfying the
# original 25-km geographical rule
#
# IMPORTANT:
# Eligibility is determined by geographical intersection
# with the 25-km Core buffer.
# Centroid distance is NOT used here.

```

```

#-----
second_ring_core_links <- lapply(
  seq_len(nrow(second_ring_sf)),
  function(i) {
    core_matches <- st_intersects(
      second_ring_sf[i, ],
      core_buffers_25km,
      sparse = TRUE
    )[[1]]
    if (length(core_matches) == 0) {
      return(NULL)
    }
    tibble(
      Second_Ring_Municipality_ID = second_ring_sf$Municipality_ID[i],
      Core_Municipality_ID = core_sf$Municipality_ID[core_matches],
      Core_Cluster_ID = core_sf$Core_Cluster_ID[core_matches]
    )
  }
) %>%
  bind_rows()
#-----
# STEP 11G: Check whether every Second Ring municipality
# has at least one qualifying Core municipality
#-----
second_ring_ids %>%
  anti_join(
    second_ring_core_links %>%
      distinct(Second_Ring_Municipality_ID),
    by = c(
      "Municipality_ID" =
        "Second_Ring_Municipality_ID"
    )
  )
#-----
# STEP 11H: Count qualifying Core municipalities and
# Core clusters for each Second Ring municipality
#-----
second_ring_cluster_candidates <- second_ring_core_links %>%
  group_by(
    Second_Ring_Municipality_ID
  ) %>%
  summarise(
    n_qualifying_core =
      n_distinct(Core_Municipality_ID),
    n_qualifying_clusters =
      n_distinct(Core_Cluster_ID),
    .groups = "drop"
  )
#-----
# STEP 11I: Inspect distribution of qualifying Core
# municipalities and Core clusters
#-----
second_ring_cluster_candidates %>%
  count(
    n_qualifying_core,
    n_qualifying_clusters,
    sort = TRUE
  )
#####
# STEP 11J: Assign Second Ring Municipalities to
# Core Clusters Using the Approved Tie-Breaker
#####
#-----
# STEP 11J-1: Create centroids
#-----

```

```

second_ring_centroids <- st_centroid(second_ring_sf)
#-----
# STEP 11J-2: Calculate distance only for qualifying
# Second Ring-Core pairs
#-----
second_ring_core_distances <- second_ring_core_links %>%
  rowwise() %>%
  mutate(
    Distance_m = as.numeric(
      st_distance(
        second_ring_centroids[
          which(
            second_ring_centroids$Municipality_ID ==
              Second_Ring_Municipality_ID
          ),
        ],
        core_centroids[
          which(
            core_centroids$Municipality_ID ==
              Core_Municipality_ID
          ),
        ]
      )
    ) %>%
  ungroup()
#-----
# STEP 11J-3: Select the nearest qualifying Core
#
# Distance is used ONLY as a tie-breaker among Core
# municipalities already satisfying the 25-km rule.
#
# Deterministic ordering:
# 1. Distance
# 2. Core_Cluster_ID
# 3. Core_Municipality_ID
#-----
second_ring_cluster_assignment <- second_ring_core_distances %>%
  arrange(
    Second_Ring_Municipality_ID,
    Distance_m,
    Core_Cluster_ID,
    Core_Municipality_ID
  ) %>%
  group_by(
    Second_Ring_Municipality_ID
  ) %>%
  slice(1) %>%
  ungroup() %>%
  select(
    Second_Ring_Municipality_ID,
    Core_Municipality_ID,
    Core_Cluster_ID,
    Distance_m
  )
#-----
# STEP 11J-4: Check one assignment per Second Ring
#-----
second_ring_cluster_assignment %>%
  count(Second_Ring_Municipality_ID) %>%
  filter(n > 1)
#-----
# STEP 11J-5: Verify all 1,171 Second Ring municipalities
# were assigned
#-----

```

```

nrow(second_ring_cluster_assignment)
n_distinct(
  second_ring_cluster_assignment$Second_Ring_Municipality_ID
)
#-----
# STEP 11J-6: Check for missing Core-cluster assignments
#-----
sum(
  is.na(
    second_ring_cluster_assignment$Core_Cluster_ID
  )
)
#####
# STEP 11K: Check for Exact Distance Ties
#####
second_ring_core_distances %>%
  group_by(Second_Ring_Municipality_ID) %>%
  summarise(
    Minimum_Distance = min(Distance_m),
    N_at_Minimum = sum(
      abs(Distance_m - min(Distance_m)) < 1e-6
    ),
    .groups = "drop"
  ) %>%
  filter(N_at_Minimum > 1)
#####
# STEP 12: Combine and Validate Core-Cluster Assignments
#####
#-----
# STEP 12A: Create Core municipality lookup
#-----
core_lookup <- core_sf %>%
  st_drop_geometry() %>%
  select(
    Municipality_ID,
    Municipality,
    Core_Cluster_ID
  ) %>%
  mutate(
    Spatial_Group = "Core"
  )
#-----
# STEP 12B: Create Neighbor municipality lookup
#-----
neighbor_lookup <- neighbor_cluster_assignment %>%
  left_join(
    neighbor_ids,
    by = c(
      "Neighbor_Municipality_ID" =
        "Municipality_ID"
    )
  ) %>%
  transmute(
    Municipality_ID =
      Neighbor_Municipality_ID,
    Municipality,
    Core_Cluster_ID,
    Spatial_Group = "Neighbor"
  )
#-----
# STEP 12C: Create Second Ring municipality lookup
#-----
second_ring_lookup <- second_ring_cluster_assignment %>%
  left_join(
    second_ring_ids,

```

```

by = c(
  "Second_Ring_Municipality_ID" =
  "Municipality_ID"
)
)%>%
transmute(
  Municipality_ID =
    Second_Ring_Municipality_ID,
  Municipality,
  Core_Cluster_ID,
  Spatial_Group = "Second Ring"
)
#-----
# STEP 12D: Combine all three spatial groups
#-----
core_cluster_lookup <- bind_rows(
  core_lookup,
  neighbor_lookup,
  second_ring_lookup
)
#-----
# STEP 12E: Check total number of municipalities
#-----
nrow(core_cluster_lookup)
n_distinct(
  core_cluster_lookup$Municipality_ID
)
#-----
# STEP 12F: Check spatial-group counts
#-----
core_cluster_lookup %>%
  count(Spatial_Group)
#-----
# STEP 12G: Check for duplicate municipality IDs
#-----
core_cluster_lookup %>%
  count(Municipality_ID) %>%
  filter(n > 1)
#-----
# STEP 12H: Check for missing Core-cluster assignments
#-----
sum(
  is.na(
    core_cluster_lookup$Core_Cluster_ID
  )
)
#-----
# STEP 12I: Check number of Core clusters represented
#-----
n_distinct(
  core_cluster_lookup$Core_Cluster_ID
)
#-----
# STEP 12J: Check municipalities by Core cluster
#-----
core_cluster_lookup %>%
  count(
    Core_Cluster_ID,
    Spatial_Group
  ) %>%
  arrange(
    Core_Cluster_ID,
    Spatial_Group
  )
#-----

```

```

# STEP 12K: Final structural validation
#-----
stopifnot(
  nrow(core_cluster_lookup) == 1814,
  n_distinct(core_cluster_lookup$Municipality_ID) == 1814,
  sum(is.na(core_cluster_lookup$Core_Cluster_ID)) == 0,
  n_distinct(core_cluster_lookup$Core_Cluster_ID) == 81
)
#####
# STEP 13: Save and Verify Final Core-Cluster Lookup
#####
#-----
# STEP 13A: Create output directory if necessary
#-----
dir.create(
  "03_Output/Data",
  recursive = TRUE,
  showWarnings = FALSE
)
dir.create(
  "03_Output/Audit",
  recursive = TRUE,
  showWarnings = FALSE
)
#-----
# STEP 13B: Save final Core-cluster lookup as RData
#-----
save(
  core_cluster_lookup,
  file = "03_Output/Data/Core_Cluster_Lookup_Final.RData"
)
#-----
# STEP 13C: Save final Core-cluster lookup as CSV
#-----
write.csv(
  core_cluster_lookup,
  "03_Output/Data/Core_Cluster_Lookup_Final.csv",
  row.names = FALSE
)
#-----
# STEP 13D: Create assignment audit table
#-----
core_cluster_assignment_audit <- core_cluster_lookup %>%
  count(
    Core_Cluster_ID,
    Spatial_Group,
    name = "Municipality_Count"
  ) %>%
  arrange(
    Core_Cluster_ID,
    Spatial_Group
  )
#-----
# STEP 13E: Save assignment audit
#-----
write.csv(
  core_cluster_assignment_audit,
  "03_Output/Audit/Core_Cluster_Assignment_Audit.csv",
  row.names = FALSE
)
#-----
# STEP 13F: Final verification after saving
#-----
file.exists(
  "03_Output/Data/Core_Cluster_Lookup_Final.RData"
)

```

```

)
file.exists(
  "03_Output/Data/Core_Cluster_Lookup_Final.csv"
)
file.exists(
  "03_Output/Audit/Core_Cluster_Assignment_Audit.csv"
)
#-----
# STEP 13G: Verify saved lookup can be reloaded
#-----
rm(core_cluster_lookup)
load(
  "03_Output/Data/Core_Cluster_Lookup_Final.RData"
)
#-----
# STEP 13H: Verify reloaded object
#-----
nrow(core_cluster_lookup)
n_distinct(
  core_cluster_lookup$Municipality_ID
)
core_cluster_lookup %>%
  count(Spatial_Group)
sum(
  is.na(
    core_cluster_lookup$Core_Cluster_ID
  )
)
n_distinct(
  core_cluster_lookup$Core_Cluster_ID
)
#-----
# STEP 13I: Final structural checks
#-----
stopifnot(
  nrow(core_cluster_lookup) == 1814,
  n_distinct(core_cluster_lookup$Municipality_ID) == 1814,
  sum(is.na(core_cluster_lookup$Core_Cluster_ID)) == 0,
  n_distinct(core_cluster_lookup$Core_Cluster_ID) == 81
)
#####
# END OF STEP 13
#####

A.2.4.2 Core-Cluster Regression Model
#####
# 04_04_Core_Cluster_Spatial_Association_Final.R
# Purpose:
# Estimate the final Core-cluster spatial association model
#
# Final specification:
# - Neighbor + Second Ring municipalities only
# - Second Ring = reference group
# - Municipality fixed effects
# - Year fixed effects
# - Core-cluster outcome
# - Neighbor × CoreOut interaction
# - Controls: Population, Beds, NightsPerResident, Foreign_Tourist_Share
# - Standard errors clustered at Core-cluster level
# using CR2 with Satterthwaite degrees of freedom
#####
# STEP 1: Load Required Packages
#####
library(dplyr)

```

```

library(tidyr)
library(plm)
library(lmtest)
library(sandwich)
library(cluster)
library(clubSandwich)
#####
# STEP 2: Load Balanced Panel
#####
setwd(
  "D:/Edu/Masters/masters thesis/Master_Thesis_V3")
load(
  "01_Data/Final_Data/Master_Analysis_Panel_Final_Spatial_2017_2024.RData")
# Verify object
stopifnot(exists("balanced_panel"))
#####
# STEP 3: Load Final Core-Cluster Lookup
#####
load("03_Output/Data/Core_Cluster_Lookup_Final.RData")
# Verify object
stopifnot(exists("core_cluster_lookup"))
#####
# STEP 4: Basic Panel Verification
#####
nrow(balanced_panel)
n_distinct(
  balanced_panel$Municipality_ID)
range(
  balanced_panel$Year,
  na.rm = TRUE)
balanced_panel %>%
  count(Year)
#####
# STEP 5: Verify Final Core-Cluster Lookup
#####
nrow(core_cluster_lookup)
n_distinct(
  core_cluster_lookup$Municipality_ID)
core_cluster_lookup %>%
  count(Spatial_Group)
n_distinct( core_cluster_lookup$Core_Cluster_ID)
#####
# DIAGNOSTIC: Check Spatial Group Variables Before Merge
#####
# Check all variables in the balanced panel
names(balanced_panel)
# Check whether Spatial_Group already exists
"Spatial_Group" %in% names(balanced_panel)
# Check whether Core_Cluster_ID already exists
"Core_Cluster_ID" %in% names(balanced_panel)
# Check the variables in the final lookup
names(core_cluster_lookup)
# Check the first few rows of the lookup
head( core_cluster_lookup)
# Check the structure of the relevant variables
str( balanced_panel$Municipality_ID)
str( core_cluster_lookup$Municipality_ID)
#####
# STEP 6: Merge Final Core-Cluster IDs into Panel
#####
core_cluster_panel <- balanced_panel %>%
  left_join( core_cluster_lookup %>%
    select( Municipality_ID, Core_Cluster_ID), by = "Municipality_ID")
#####
# STEP 7: Verify Merge
#####

```

```

nrow(core_cluster_panel)
n_distinct(
  core_cluster_panel$Municipality_ID)
sum(
  is.na(
    core_cluster_panel$Core_Cluster_ID))
#####
# STEP 8: Verify Existing Spatial Groups
#####
core_cluster_panel %>%
  distinct(
    Municipality_ID,
    Spatial_Group
  ) %>%
  count(Spatial_Group)
#####
# STEP 9: Check Cluster IDs by Spatial Group
#####
core_cluster_panel %>%
  filter(
    Spatial_Group %in% c(
      "Core",
      "Neighbor",
      "Second Ring"
    )
  ) %>%
  distinct(
    Municipality_ID,
    Spatial_Group,
    Core_Cluster_ID
  ) %>%
  count(
    Spatial_Group
  )
#####
# STEP 10: Verify Other Municipalities
#####
core_cluster_panel %>%
  filter(
    Spatial_Group == "Other"
  ) %>%
  summarise(
    Municipalities = n_distinct(Municipality_ID),
    Missing_Cluster_ID = sum(
      is.na(Core_Cluster_ID)
    )
  )
#####
# STEP 11: Final Merge Validation
#####
stopifnot(
  nrow(core_cluster_panel) == 21944,
  n_distinct(core_cluster_panel$Municipality_ID) == 2743,
  sum(
    is.na(
      core_cluster_panel$Core_Cluster_ID
    )
  ) == 7432,
  n_distinct(
    core_cluster_panel$Core_Cluster_ID[
      !is.na(core_cluster_panel$Core_Cluster_ID)
    ]
  ) == 81
)
#####

```

```

# END OF MERGE VALIDATION
#####
#####
# STEP 12: Create Final Core-Cluster Estimation Sample
#####
core_cluster_model_data <- core_cluster_panel %>%
  filter(
    Spatial_Group %in% c(
      "Neighbor",
      "Second Ring"
    )
  )
#####
# STEP 13: Verify Estimation-Sample Structure
#####
core_cluster_model_data %>%
  distinct(
    Municipality_ID,
    Spatial_Group
  ) %>%
  count(Spatial_Group)
n_distinct(
  core_cluster_model_data$Municipality_ID
)
n_distinct(
  core_cluster_model_data$Core_Cluster_ID
)
range(
  core_cluster_model_data$Year,
  na.rm = TRUE
)
#####
# STEP 14: Verify Municipality-Year Coverage
#####
core_cluster_model_data %>%
  count(
    Municipality_ID
  ) %>%
  summarise(
    Minimum_Years = min(n),
    Maximum_Years = max(n),
    Mean_Years = mean(n)
  )
#####
# STEP 15: Verify Observations by Year and Group
#####
core_cluster_model_data %>%
  count(
    Year,
    Spatial_Group
  ) %>%
  arrange(
    Year,
    Spatial_Group
  )
#####
# STEP 16: Check Core-Cluster Coverage
#####
core_cluster_model_data %>%
  distinct(
    Core_Cluster_ID
  ) %>%
  summarise(
    Number_of_Core_Clusters =
      n_distinct(Core_Cluster_ID))

```

```
#####
# STEP 17: Check Required Model Variables
#####
required_variables <- c(
  "Municipality_ID",
  "Year",
  "Spatial_Group",
  "Core_Cluster_ID",
  "TouristTaxRevenue",
  "TouristTaxPerResident",
  "TouristTaxPerNight",
  "Population",
  "Beds",
  "NightsPerResident",
  "Foreign_Tourist_Share")

setdiff(
  required_variables,
  names(core_cluster_model_data))
#####
# STEP 18: Structural Validation
#####
stopifnot(
  n_distinct(
    core_cluster_model_data$Municipality_ID
  ) == 1652,
  n_distinct(
    core_cluster_model_data$Core_Cluster_ID
  ) == 81,
  all(
    core_cluster_model_data$Spatial_Group %in%
    c(
      "Neighbor",
      "Second Ring"
    )
  ),
  all(
    core_cluster_model_data %>%
    count(Municipality_ID) %>%
    pull(n) == 8
  )
)
#####
# END OF ESTIMATION-SAMPLE VALIDATION
#####
# STEP 19: Construct CoreOut at Core-Cluster-Year Level
#####
#-----
# STEP 19A: Extract Core Municipalities from Final Panel
#-----
core_municipality_panel <- core_cluster_panel %>%
  filter(
    Spatial_Group == "Core"
  ) %>%
  select(
    Municipality_ID,
    Year,
    Core_Cluster_ID,
    TouristTaxPerResident,
    TouristTaxPerNight
  )
#-----
# STEP 19B: Verify Core Municipality-Year Data
#-----
```

```

n_distinct(
  core_municipality_panel$Municipality_ID)
n_distinct(
  core_municipality_panel$Core_Cluster_ID)
range(
  core_municipality_panel$Year,
  na.rm = TRUE)
#####
# STEP 19C: Construct Core-Cluster-Year CoreOut Measures
#####
# CoreOut is defined as the unweighted arithmetic mean of
# the corresponding outcome across Core municipalities
# belonging to each Core cluster in each year.
#
# For the Tourist Tax per Resident model:
# CoreOut_TaxPerResident =
# mean of TouristTaxPerResident across Core municipalities
# within the same Core cluster and year.
#
# For the Tourist Tax per Night model:
# CoreOut_TaxPerNight =
# mean of TouristTaxPerNight across Core municipalities
# within the same Core cluster and year.
#
# Each Core municipality receives equal weight.
# No population weighting or ratio of aggregate totals is used.
#
# This produces one observation for each
# Core cluster-year combination.
core_out <- core_municipality_panel %>% group_by(Core_Cluster_ID, Year) %>% summarise(
  CoreOut_TaxPerResident = mean(TouristTaxPerResident, na.rm = TRUE),
  CoreOut_TaxPerNight = mean(TouristTaxPerNight, na.rm = TRUE),
  .groups = "drop"
)
#-----
# STEP 19D: Verify CoreOut Structure
#-----
nrow(core_out)
n_distinct(core_out$Core_Cluster_ID)
n_distinct(core_out$Year)
summary(core_out$CoreOut_TaxPerResident)
summary(core_out$CoreOut_TaxPerNight)
#-----
# STEP 19E: Check One Observation per Core Cluster-Year
#-----
core_out %>% count(Core_Cluster_ID, Year) %>% filter(n > 1)
#-----
# STEP 19F: Check Missing and Invalid CoreOut Values
#-----
sum(is.na(core_out$CoreOut_TaxPerResident))
sum(is.na(core_out$CoreOut_TaxPerNight))
sum(!is.finite(core_out$CoreOut_TaxPerResident))
sum(!is.finite(core_out$CoreOut_TaxPerNight))
#-----
# STEP 19G: Verify Expected 81 × 8 Structure
#-----
stopifnot(
  nrow(core_out) == 648,
  n_distinct(core_out$Core_Cluster_ID) == 81,
  n_distinct(core_out$Year) == 8,
  sum(is.na(core_out$CoreOut_TaxPerResident)) == 0,
  sum(is.na(core_out$CoreOut_TaxPerNight)) == 0,
  sum(!is.finite(core_out$CoreOut_TaxPerResident)) == 0,
  sum(!is.finite(core_out$CoreOut_TaxPerNight)) == 0
)

```

```
#####
# END OF STEP 19
#####
#####
# STEP 20: Attach CoreOut Measures to Estimation Sample
#####
core_cluster_model_data <- core_cluster_model_data %>% left_join(core_out %>% select(Core_Cluster_ID, Year,
CoreOut_TaxPerResident, CoreOut_TaxPerNight), by = c("Core_Cluster_ID", "Year"))
#####
# STEP 21: Verify CoreOut Merge
#####
nrow(core_cluster_model_data)
n_distinct(core_cluster_model_data$Municipality_ID)
sum(is.na(core_cluster_model_data$CoreOut_TaxPerResident))
sum(is.na(core_cluster_model_data$CoreOut_TaxPerNight))
#####
# STEP 22: Check CoreOut Measures by Spatial Group
#####
core_cluster_model_data %>% group_by(Spatial_Group) %>% summarise(
  Municipalities = n_distinct(Municipality_ID),
  Observations = n(),
  Missing_CoreOut_TaxPerResident = sum(is.na(CoreOut_TaxPerResident)),
  Missing_CoreOut_TaxPerNight = sum(is.na(CoreOut_TaxPerNight)),
  Min_CoreOut_TaxPerResident = min(CoreOut_TaxPerResident, na.rm = TRUE),
  Max_CoreOut_TaxPerResident = max(CoreOut_TaxPerResident, na.rm = TRUE),
  Min_CoreOut_TaxPerNight = min(CoreOut_TaxPerNight, na.rm = TRUE),
  Max_CoreOut_TaxPerNight = max(CoreOut_TaxPerNight, na.rm = TRUE),
  .groups = "drop"
)
#####
# STEP 23: Verify One CoreOut Value per Core Cluster-Year
#####
core_cluster_model_data %>% distinct(Core_Cluster_ID, Year, CoreOut_TaxPerResident, CoreOut_TaxPerNight) %>%
count(Core_Cluster_ID, Year) %>% filter(n > 1)
#####
# STEP 24: Create Neighbor × CoreOut Interaction Variables
#####
core_cluster_model_data <- core_cluster_model_data %>% mutate(
  Neighbor = ifelse(Spatial_Group == "Neighbor", 1, 0),
  Neighbor_CoreOut_TaxPerResident = Neighbor * CoreOut_TaxPerResident,
  Neighbor_CoreOut_TaxPerNight = Neighbor * CoreOut_TaxPerNight
)
#####
# STEP 25: Verify Interaction Variables
#####
core_cluster_model_data %>% count(Spatial_Group, Neighbor)
summary(core_cluster_model_data$CoreOut_TaxPerResident)
summary(core_cluster_model_data$CoreOut_TaxPerNight)
summary(core_cluster_model_data$Neighbor_CoreOut_TaxPerResident)
summary(core_cluster_model_data$Neighbor_CoreOut_TaxPerNight)
#####
# STEP 26: Final Validation Before Regression
#####
stopifnot(
  nrow(core_cluster_model_data) == 13216,
  n_distinct(core_cluster_model_data$Municipality_ID) == 1652,
  n_distinct(core_cluster_model_data$Core_Cluster_ID) == 81,
  sum(is.na(core_cluster_model_data$CoreOut_TaxPerResident)) == 0,
  sum(is.na(core_cluster_model_data$CoreOut_TaxPerNight)) == 0,
  all(core_cluster_model_data$Neighbor %in% c(0, 1))
)
#####
# END OF STEP 26
#####
#####
```

```

# STEP 27: Check Final Core-Cluster Model Variables
#####
# Final model controls
core_model_controls <- c(
  "Population",
  "Beds",
  "NightsPerResident",
  "Foreign_Tourist_Share"
)
# Final Core-cluster outcomes
core_model_outcomes <- c(
  "TouristTaxPerResident",
  "TouristTaxPerNight"
)
#####
# STEP 28: Check Missingness in Model Variables
#####
core_cluster_model_data %>%
  summarise(
    Missing_TouristTaxPerResident =
      sum(is.na(TouristTaxPerResident)),
    Missing_TouristTaxPerNight =
      sum(is.na(TouristTaxPerNight)),
    Missing_Population =
      sum(is.na(Population)),
    Missing_Beds =
      sum(is.na(Beds)),
    Missing_NightsPerResident =
      sum(is.na(NightsPerResident)),
    Missing_Foreign_Tourist_Share =
      sum(is.na(Foreign_Tourist_Share))
  )
#####
# STEP 29: Check Completeness of Each Model
#####
core_cluster_model_data %>%
  summarise(
    Municipalities =
      n_distinct(Municipality_ID),
    Observations =
      n(),
    Complete_Resident_Model =
      sum(
        complete.cases(
          select(
            `
            TouristTaxPerResident,
            Population,
            Beds,
            NightsPerResident,
            Foreign_Tourist_Share,
            CoreOut_TaxPerResident,
            Neighbor_CoreOut_TaxPerResident
          )
        )
      ),
    Complete_Night_Model =
      sum(
        complete.cases(
          select(
            `
            TouristTaxPerNight,
            Population,
            Beds,
            NightsPerResident,

```

```

        Foreign_Tourist_Share,
        CoreOut_TaxPerNight,
        Neighbor_CoreOut_TaxPerNight
    )
)
)
)
)
#####
# STEP 30: Identify Missing Observations by Model
#####
#-----
# Resident Model
#-----
core_cluster_model_data %>%
  filter(
    !complete.cases(
      select(
        `
        TouristTaxPerResident,
        Population,
        Beds,
        NightsPerResident,
        Foreign_Tourist_Share,
        CoreOut_TaxPerResident,
        Neighbor_CoreOut_TaxPerResident
      )
    )
  ) %>%
  count(
    Municipality_ID,
    Spatial_Group
  )
#-----
# Night Model
#-----
core_cluster_model_data %>%
  filter(
    !complete.cases(
      select(
        `
        TouristTaxPerNight,
        Population,
        Beds,
        NightsPerResident,
        Foreign_Tourist_Share,
        CoreOut_TaxPerNight,
        Neighbor_CoreOut_TaxPerNight
      )
    )
  ) %>%
  count(
    Municipality_ID,
    Spatial_Group
  )
#####
# STEP 31: Check ExtraHotel_Share Separately
#####
core_cluster_model_data %>%
  summarise(
    Missing_ExtraHotel_Share =
      sum(is.na(ExtraHotel_Share)),
    Total_Observations =
      n(),
    Missing_Percent =
      100 *

```

```

    mean(is.na(ExtraHotel_Share))
  )
#####
# STEP 32: Verify Final Outcome and CoreOut Variables
#####
setdiff(
  c(
    core_model_outcomes,
    core_model_controls,
    "CoreOut_TaxPerResident",
    "CoreOut_TaxPerNight",
    "Neighbor",
    "Neighbor_CoreOut_TaxPerResident",
    "Neighbor_CoreOut_TaxPerNight"
  ),
  names(core_cluster_model_data)
)
#####
# END OF VARIABLE AND MISSINGNESS CHECK
#####
#####
# DIAGNOSTIC: Identify Missing Variable for Municipality 001284
#####
core_cluster_model_data %>%
  filter(
    Municipality_ID == "001284"
  ) %>%
  select(
    Municipality_ID,
    Year,
    Spatial_Group,
    TouristTaxPerResident,
    TouristTaxPerNight,
    Population,
    Beds,
    NightsPerResident,
    Foreign_Tourist_Share,
    CoreOut_TaxPerResident,
    CoreOut_TaxPerNight,
    Neighbor_CoreOut_TaxPerResident,
    Neighbor_CoreOut_TaxPerNight
  ) %>%
  filter(
    if_any(
      everything(),
      is.na
    )
  )
#####
# DIAGNOSTIC: Identify Exact Missing Variable
#####
core_cluster_model_data %>%
  filter(
    Municipality_ID == "001284",
    Year == 2019
  ) %>%
  summarise(
    Foreign_Tourist_Share_Missing =
      is.na(Foreign_Tourist_Share),
    CoreOut_TaxPerResident_Missing =
      is.na(CoreOut_TaxPerResident),
    CoreOut_TaxPerNight_Missing =
      is.na(CoreOut_TaxPerNight),
    Neighbor_CoreOut_TaxPerResident_Missing =
      is.na(Neighbor_CoreOut_TaxPerResident),

```

```

Neighbor_CoreOut_TaxPerNight_Missing =
  is.na(Neighbor_CoreOut_TaxPerNight)
)
#####
# STEP 33: Inspect Final Core-Cluster Model Specification
#####
#-----
# Check variable classes
#-----
str(
  core_cluster_model_data %>%
    select(
      Municipality_ID,
      Year,
      Spatial_Group,
      Core_Cluster_ID,
      CoreOut_TaxPerResident,
      CoreOut_TaxPerNight,
      Neighbor,
      Neighbor_CoreOut_TaxPerResident,
      Neighbor_CoreOut_TaxPerNight,
      TouristTaxPerResident,
      TouristTaxPerNight,
      Population,
      Beds,
      NightsPerResident,
      Foreign_Tourist_Share
    )
)
#-----
# Check spatial-group coding
#-----
core_cluster_model_data %>%
  distinct(
    Municipality_ID,
    Spatial_Group,
    Neighbor
  ) %>%
  count(
    Spatial_Group,
    Neighbor
  )
#-----
# Check CoreOut variation over time
#-----
core_cluster_model_data %>%
  group_by(
    Core_Cluster_ID
  ) %>%
  summarise(
    Years =
      n_distinct(Year),
    CoreOut_TaxPerResident_Min =
      min(
        CoreOut_TaxPerResident
      ),
    CoreOut_TaxPerResident_Max =
      max(
        CoreOut_TaxPerResident
      ),
    CoreOut_TaxPerResident_SD =
      sd(
        CoreOut_TaxPerResident
      ),
    CoreOut_TaxPerNight_Min =

```

```

    min(
      CoreOut_TaxPerNight
    ),
    CoreOut_TaxPerNight_Max =
      max(
        CoreOut_TaxPerNight
      ),
    CoreOut_TaxPerNight_SD =
      sd(
        CoreOut_TaxPerNight
      ),
    .groups = "drop"
  ) %>%
  summarise(
    Minimum_Years =
      min(Years),
    Maximum_Years =
      max(Years),
    Clusters_With_Time_Variation_TaxPerResident =
      sum(
        CoreOut_TaxPerResident_SD > 0
      ),
    Clusters_With_Time_Variation_TaxPerNight =
      sum(
        CoreOut_TaxPerNight_SD > 0
      )
  )
#-----
# Check Neighbor interaction mathematically
#-----
all.equal(
  core_cluster_model_data$Neighbor_CoreOut_TaxPerResident,
  core_cluster_model_data$Neighbor *
  core_cluster_model_data$CoreOut_TaxPerResident
)
all.equal(
  core_cluster_model_data$Neighbor_CoreOut_TaxPerNight,
  core_cluster_model_data$Neighbor *
  core_cluster_model_data$CoreOut_TaxPerNight
)
#-----
# Check number of municipalities per Core cluster
#-----
core_cluster_model_data %>%
  distinct(
    Municipality_ID,
    Spatial_Group,
    Core_Cluster_ID
  ) %>%
  count(
    Core_Cluster_ID,
    Spatial_Group
  ) %>%
  arrange(
    Core_Cluster_ID,
    Spatial_Group
  )
#-----
# Check final complete-case samples
#-----
resident_complete <- core_cluster_model_data %>%
  filter(
    complete.cases(
      TouristTaxPerResident,
      CoreOut_TaxPerResident,

```

```

    Neighbor,
    Population,
    Beds,
    NightsPerResident,
    Foreign_Tourist_Share
  )
)
night_complete <- core_cluster_model_data %>%
  filter(
    complete.cases(
      TouristTaxPerNight,
      CoreOut_TaxPerNight,
      Neighbor,
      Population,
      Beds,
      NightsPerResident,
      Foreign_Tourist_Share
    )
  )
#-----
# Complete-case sample sizes
#-----
nrow(resident_complete)
n_distinct(
  resident_complete$Municipality_ID
)
n_distinct(
  resident_complete$Core_Cluster_ID
)
nrow(night_complete)
n_distinct(
  night_complete$Municipality_ID
)
n_distinct(
  night_complete$Core_Cluster_ID
)
#####
# END OF STEP 33
#####
#####
# STEP 34: Core-c-Cluster Model
# Outcome: Tourist Tax per Resident
#####
resident_panel <- pdata.frame(resident_complete, index = c("Municipality_ID", "Year"))
core_model_resident <- plm(TouristTaxPerResident ~ CoreOut_TaxPerResident + Neighbor_CoreOut_TaxPerResident +
  Population + Beds + NightsPerResident + Foreign_Tourist_Share + factor(Year), data = resident_panel, model = "within",
  effect = "individual")
summary(core_model_resident)
#####
# END OF STEP 34
#####
#####
# STEP 35: Core-Cluster CR2 Inference
# Outcome: Tourist Tax per Resident
#####
core_resident_cr2 <- vcovCR(core_model_resident, cluster = resident_complete$Core_Cluster_ID, type = "CR2")
core_resident_tests <- coef_test(core_model_resident, vcov = core_resident_cr2, test = "Satterthwaite")
core_resident_tests
#####
# END OF STEP 35
#####
#####
# STEP 36: Linear Combination Test
# Neighbor Effect =  $\beta + \delta$ 
#####

```

```

names(coef(core_model_resident))
resident_contrast <- matrix(0, nrow = 1, ncol = length(coef(core_model_resident)))
colnames(resident_contrast) <- names(coef(core_model_resident))
resident_contrast[1, "CoreOut_TaxPerResident"] <- 1
resident_contrast[1, "Neighbor_CoreOut_TaxPerResident"] <- 1
core_resident_neighbor_effect <- linear_contrast(core_model_resident, contrasts = resident_contrast, vcov =
core_resident_cr2, test = "Satterthwaite")
core_resident_neighbor_effect
#####
# END OF STEP 36
#####
#####
# STEP 37: Core-Cluster Model
# Outcome: Tourist Tax per Night
#####
night_panel <- pdata.frame(night_complete, index = c("Municipality_ID", "Year"))
core_model_night <- plm(TouristTaxPerNight ~ CoreOut_TaxPerNight + Neighbor_CoreOut_TaxPerNight + Population +
Beds + NightsPerResident + Foreign_Tourist_Share + factor(Year), data = night_panel, model = "within", effect =
"individual")
summary(core_model_night)
#####
# END OF STEP 37
#####
#####
# STEP 38: Core-Cluster CR2 Standard Errors
# Outcome: Tourist Tax per Night
#####
core_night_cr2 <- vcovCR(core_model_night, cluster = night_complete$Core_Cluster_ID, type = "CR2")
core_night_tests <- coef_test(core_model_night, vcov = core_night_cr2, test = "Satterthwaite")
core_night_tests
#####
# END OF STEP 38
#####
#####
# STEP 39: Linear Contrast ( $\beta + \delta$ )
# Outcome: Tourist Tax per Night
#####
names(coef(core_model_night))
night_contrast <- matrix(0, nrow = 1, ncol = length(coef(core_model_night)))
colnames(night_contrast) <- names(coef(core_model_night))
night_contrast[1, "CoreOut_TaxPerNight"] <- 1
night_contrast[1, "Neighbor_CoreOut_TaxPerNight"] <- 1
core_night_neighbor_effect <- linear_contrast(core_model_night, contrasts = night_contrast, vcov = core_night_cr2, test =
"Satterthwaite")
core_night_neighbor_effect
#####
# END OF STEP 39
#####
#####
# STEP 40: Final Core-Cluster Model Verification
#####
print(nobs(core_model_resident))
print(nobs(core_model_night))
print(n_distinct(resident_complete$Municipality_ID))
print(n_distinct(night_complete$Municipality_ID))
print(n_distinct(resident_complete$Core_Cluster_ID))
print(n_distinct(night_complete$Core_Cluster_ID))
print(table(resident_complete$Spatial_Group))
print(table(night_complete$Spatial_Group))
print(sort(unique(resident_complete$Year)))
print(sort(unique(night_complete$Year)))

#####
# STEP 41: Save Final Core-Cluster Model Results
#####

```

```

Core_Cluster_Final_Results <- list(
  Resident_Model = core_model_resident,
  Resident_CR2 = core_resident_cr2,
  Resident_Tests = core_resident_tests,
  Resident_Neighbor_Effect = core_resident_neighbor_effect,
  Night_Model = core_model_night,
  Night_CR2 = core_night_cr2,
  Night_Tests = core_night_tests,
  Night_Neighbor_Effect = core_night_neighbor_effect
)
save(Core_Cluster_Final_Results, file = "03_Output/Data/Core_Cluster_Final_Results.RData")
print(file.exists("03_Output/Data/Core_Cluster_Final_Results.RData"))
#####
# STEP 42: Extract Final Core-Cluster Coefficient Results
#####
resident_results_table <- as.data.frame(Core_Cluster_Final_Results$Resident_Tests)
resident_results_table$Outcome <- "Tourist Tax per Resident"
night_results_table <- as.data.frame(Core_Cluster_Final_Results$Night_Tests)
night_results_table$Outcome <- "Tourist Tax per Night"
core_cluster_results_table <- rbind(resident_results_table, night_results_table)
print(core_cluster_results_table)
#####
# STEP 43: Extract Neighbor Total Effects ( $\beta + \delta$ )
#####
resident_neighbor_effect <- as.data.frame(Core_Cluster_Final_Results$Resident_Neighbor_Effect)
resident_neighbor_effect$Outcome <- "Tourist Tax per Resident"
night_neighbor_effect <- as.data.frame(Core_Cluster_Final_Results$Night_Neighbor_Effect)
night_neighbor_effect$Outcome <- "Tourist Tax per Night"
neighbor_total_effects <- rbind(resident_neighbor_effect, night_neighbor_effect)
print(neighbor_total_effects)
#####
# STEP 44: Final Core-Cluster Central Results Table
#####
resident_beta <- resident_results_table[resident_results_table$Coef == "CoreOut_TaxPerResident",]
resident_delta <- resident_results_table[resident_results_table$Coef == "Neighbor_CoreOut_TaxPerResident",]
night_beta <- night_results_table[night_results_table$Coef == "CoreOut_TaxPerNight",]
night_delta <- night_results_table[night_results_table$Coef == "Neighbor_CoreOut_TaxPerNight",]
core_cluster_central_results <- data.frame(
  Outcome = c("Tourist Tax per Resident", "Tourist Tax per Night"),
  Beta = c(resident_beta$beta, night_beta$beta),
  Beta_SE = c(resident_beta$SE, night_beta$SE),
  Beta_p = c(resident_beta$p_Satt, night_beta$p_Satt),
  Delta = c(resident_delta$beta, night_delta$beta),
  Delta_SE = c(resident_delta$SE, night_delta$SE),
  Delta_p = c(resident_delta$p_Satt, night_delta$p_Satt),
  Beta_Plus_Delta = c(resident_neighbor_effect$Est, night_neighbor_effect$Est),
  Beta_Plus_Delta_SE = c(resident_neighbor_effect$SE, night_neighbor_effect$SE),
  Beta_Plus_Delta_Lower95 = c(resident_neighbor_effect$CI_L, night_neighbor_effect$CI_L),
  Beta_Plus_Delta_Upper95 = c(resident_neighbor_effect$CI_U, night_neighbor_effect$CI_U)
)
print(core_cluster_central_results)
#####
# END OF STEP 44
#####
# STEP 45: Add Satterthwaite Degrees of Freedom
#####
core_cluster_central_results$Beta_df <- c(resident_beta$df_Satt, night_beta$df_Satt)
core_cluster_central_results$Delta_df <- c(resident_delta$df_Satt, night_delta$df_Satt)
core_cluster_central_results$Beta_Plus_Delta_df <- c(resident_neighbor_effect$df, night_neighbor_effect$df)
print(core_cluster_central_results)
#####
# STEP 46: Save Final Core-Cluster Central Results
#####
save(core_cluster_central_results, file = "03_Output/Data/Core_Cluster_Central_Results.RData")

```

```

write.csv(core_cluster_central_results, file = "03_Output/Data/Core_Cluster_Central_Results.csv", row.names = FALSE)
print(file.exists("03_Output/Data/Core_Cluster_Central_Results.RData"))
print(file.exists("03_Output/Data/Core_Cluster_Central_Results.csv"))
#####
# STEP 47: Reload and Verify Final Central Results
#####
load("03_Output/Data/Core_Cluster_Central_Results.RData")
print(dim(core_cluster_central_results))
print(core_cluster_central_results$Outcome)
print(all.equal(core_cluster_central_results$Beta[1], core_resident_tests$beta[core_resident_tests$Coef ==
"CoreOut_TaxPerResident"])))
print(all.equal(core_cluster_central_results$Beta[2], core_night_tests$beta[core_night_tests$Coef ==
"CoreOut_TaxPerNight"])))
print(all.equal(core_cluster_central_results$Delta[1], core_resident_tests$beta[core_resident_tests$Coef ==
"Neighbor_CoreOut_TaxPerResident"])))
print(all.equal(core_cluster_central_results$Delta[2], core_night_tests$beta[core_night_tests$Coef ==
"Neighbor_CoreOut_TaxPerNight"])))
print(all.equal(core_cluster_central_results$Beta_Plus_Delta[1], core_resident_neighbor_effect$Est))
print(all.equal(core_cluster_central_results$Beta_Plus_Delta[2], core_night_neighbor_effect$Est))
#####
# END OF STEP 47
#####

```