CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Alexander Kling

The Effect of Cultural and Natural Heritage on Tourism Attractiveness in Slovakia

Bachelor's thesis

Prague 2022

Author: Alexander Kling
Supervisor: PhDr. Lenka Šťastná Ph.D.

Academic Year: 2021/2022

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, August 2, 2022

Alexander Kling

Abstract

This thesis questions a traditional understanding of cultural and natural heritage as two of the main determinants of tourism attractiveness. The motivation to write it was that in Slovakia, there are multiple locations with plentiful natural and cultural monuments, but since some of them are popular tourist destinations, the others are relatively unknown from the tourism perspective. Based on the review of the existing academic literature, the author understands the concept of attractiveness as a phenomenon composed of the attributes of a certain attraction or location and tourists‘ perception of the destinations. Therefore, the models connect both aspects of the tourism system – supply and demand for tourism. The dependent variable representing demand is the number of visitors in accommodation establishments per district in 2018 or this value per capita. The main independent variables are the number of national cultural monuments and national natural monuments and reserves, the variables typical for a tourism product. Despite the estimates of their effects seem to be clearly positive and significant, especially for nature, before including additional explanatory variables and their logarithmic transformations, in the extended models, the significant positive effect remained only for the cultural heritage, with an even higher level of confidence. The important role of a distance from the capital was found, resulting from the fact that the least developed regions with higher unemployment rates are generally located in the eastern part of the country. Some tests and models for the spatial interaction effects were also performed in the thesis, but none of them were sufficiently significant. The results of the analysis indicate that tourism policymakers should, in the case of natural heritage, focus mainly on the promotion of other destinations rich in natural beauties, not only the most visited Tatra region. In the case of cultural heritage, its positive effect was proved. However, because many districts with this type of monuments are located in east Slovakia, the investments in this region are necessary to converge more to the west of the country and fully use its potential.

Keywords: tourism, cultural heritage, natural heritage, attractiveness, regional analysis

Author's email: 37806035@fsv.cuni.cz
Supervisor's email: stastna@fsv.cuni.cz

Abstrakt

Táto práca spochybňuje tradičné chápanie kultúrneho a prírodného dedičstva ako dva z hlavných determinantov turistickej atraktivity. Motiváciou k jej napísaniu bolo to, že na Slovensku sa nachádza viacero lokalít bohatých na kultúrne a prírodne pamiatky, no kým niektoré sú populárne turistické destinácie, iné sú z hľadiska turizmu relatívne neznáme. Na základe prieskumu existujúcej literatúry chápe autor koncept atraktivity ako fenomén skladajúci sa nielen z vlastností danej atrakcie či lokality, ale aj toho, ako návštevníci destináciu vnímajú. V modeloch sa preto spájajú obidva aspekty turistického systému - ponuka a dopyt po turizme. Ako závislá premenná, zastupujúca dopyt, je použitý počet návštevníkov v ubytovacích zariadeniach v jednotlivých okresoch za rok 2018, resp. tento počet na obyvateľa. Hlavnými nezávislými premennými sú počet národných kultúrnych pamiatok a národných prírodných pamiatok a rezervácií, teda premenné typické pre turistický produkt. Kým odhady ich efektov sa zdajú byť pred pridaním ďalších vysvetľujúcich premenných a ich logaritmickou transformáciou jasne kladné a signifikantné najmä v prípade prírody, v rozšírených modeloch ostal efekt pozitívny už len pri kultúrnom dedičstve, a to význame. Dôležitú úlohu pri tom zohráva vzdialenosť od hlavného mesta, nakoľko menej rozvinuté regióny s vyššou nezamestnanosťou sa nachádzajú vo všeobecnosti viac smerom na východ. V práci boli vykonané aj niektoré testy a modely na priestorové efekty, no žiadne významne signifikantné sa nenašli. Výsledky naznačujú, že v prípade prírodného dedičstva by sa kompetentní mali zamerať na propagáciu ďalších destinácií s bohatou prírodou, nie len na najnavštevovanejší región Tatier. V prípade kultúrnych pamiatok bol dokázaný ich pozitívny efekt, no nakoľko sa množstvo okresov s týmto typom dedičstva nachádza na východnom Slovensku, pre využitie ich potenciálu sú potrebné investície do tohto regiónu, aby sa začal viac dorovnávať západu krajiny.

Klíčová slova: turismus, kulturní dědictví, přírodní dědictví, atraktivita, regionální analýza

E-mail autora 37806035@fsv.cuni.cz
E-mail vedoucího práce stastna@fsv.cuni.cz

Acknowledgments

I would like to express my gratitude to my supervisor, PhDr. Lenka Šťastná Ph.D., for her endless patience and valuable suggestions and comments. I would also like to thank Mgr. Michal Kotrč for providing me with the weighting matrices and the introduction to their codes. Last but not least, I would like to say thank you to doc. PhDr. Jozef Baruník Ph.D. for the opportunity to show him an early stage of the analysis for feedback in his course, as well as to Bc. Martina Lušková, Bc. Matěj Kořínek and Bc. Peter Palovič, for a possibility to discuss the problems I faced in the thesis with them.

Bibliographic record

Kling, Alexander: *The Effect of Cultural and Natural Heritage on Tourism Attractiveness in Slovakia.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2022, pages (113). Advisor: PhDr. Lenka Šťastná Ph.D.

Contents

List of Tables

List of Figures

Acronyms

SAC model combining the effects of SAR and SEM
SAR Spatial autoregressive model
SEM Spatial error model
SLX Spatial lag of X model
LM test Lagrangian multiplier test
RESET Regression specification error test
VIF Variance inflation factor

1 Introduction

The tourism industry belongs to the sectors hit hardest by the COVID-19 pandemics. According to the estimates published by World Tourism Organization (2021), the international tourist arrivals fell globally by 73% in the first year of the pandemics, and this drop was responsible for 42% of the total loss in international trade in 2020. The impact of losses was undoubtedly the most significant in those countries, where tourism accounts for an essential part of their total GDP.

To illustrate how crucial tourism is at a global level, we can look at some WTTC - World Travel & Tourism Council (2020) statistics before the pandemics. In 2019 travel and tourism generated directly, indirectly, and by induced impact 10.3% of global GDP and was responsible for 1 in 10 jobs worldwide. Furthermore, between 2015 and 2019, approximately a quarter of all new jobs in the world were generated in this industry. Travel & Tourism announced the highest GDP growth after Information & Communication and Financial services in 2019. All of this made this sector one of the world's largest and fastest-growing before the pandemics. (WTTC - World Travel & Tourism Council 2020)

Also, tourism in Slovakia was heavily affected by pandemic waves and lockdowns. Already the first pandemic year was the most difficult one for the travel agencies in the country in their history, according to the Statistical Office of the Slovak Republic (2021). Organized inbound tourism should decrease by almost 93% between 2019 and 2020. On the other hand, the limited possibility to travel outside the borders probably forced some travelers to consider holidays in their home countries. The drop in the number of domestic tourists using the services of travel agencies was smaller, around 32%. (Statistical Office of the Slovak Republic 2021)

Considering that foreign visitors in Slovakia usually represent a minority of all tourists (Statistical Office of the Slovak Republic n.d. *a*), we could remember some interesting situations from the pandemic time. During the peak of summer 2020, the number of visitors in one day broke a 40-year-long record. The news agencies provided videos of some parts of the Tatra mountains, which were so crowded with visitors that the tourists had to wait in queues to get to peaks (SITA 2020). The data also show an interesting observation. Even though the number of visitors in accommodation establishments in August 2020 (the most successful month by the number of tourists and also a month with relatively loose pandemic restrictions) decreased by around 13.6% compared to the previous year at the level of the whole of Slovakia, the region of Žilina recorded around 3.6% increase (Statistical Office of the Slovak Republic n.d. *b*). This region covers multiple national parks, including wide areas of the Tatra mountains.

The High and Low Tatras region belongs to the most popular tourist destinations in Slovakia (Statistical Office of the Slovak Republic 2018 *c*). There is no doubt that tourists are attracted there, among other things, by spectacular natural sceneries and hiking opportunities. The answers to whether the natural heritage sites have an important attracting power might thus look obvious. However, Slovakia has nine natural parks, some of which are located in the regions with the lowest numbers of tourism arrivals (Statistical Office of the Slovak Republic 2018*c*). A similar observation can also be found for the cultural heritage.

In order to set effective support for regional tourism development, it is crucial to start by stating what the most attractive locations are and why different districts with similar amounts of cultural and natural beauties have such different numbers of tourists. Answering those questions might provide useful information on which sectors the local authorities should invest in first to use the tourism potential of a specific destination maximally and with regard to sustainable development.

In this thesis, we will focus on some of the most critical factors in making a particular location an attractive tourist destination: natural and cultural heritage. We will examine how important they are in the context of tourism in Slovakia and how the beauties of specific regions influence the neighboring areas. Finally, we will try to find some differential characteristics of Slovak tourism across districts with a similar number of monuments but a different number of visitors.

2 Theoretical background and literature review

With the increasing importance of tourism, we can observe the roots of a study that examines this phenomenon. An early theory of tourism was introduced already in the 1920s, but since the 1960s, a new field of tourism studies has emerged and replaced the old theory. This was the era when the last phase of tourism development, as we know it today, began. Tourism became cheaper than before, new destinations and types of holidays were set in, and traveling grew into more accessible activity for broader masses. (Gyr 2010)

A booming travel and tourism industry turned out to be a relevant topic also for economists and other scientists in the second half of the 20th century. Song & Li (2008) pointed out that although in the late 80s, there were only a few academic journals oriented specifically on the research of tourism-related topics, their amount was higher than 70 around 2008.

The basic concept of studying the economics of industries is supply and demand analysis (Tisdell 2013). In the tourism industry, the supply side can be understood as a destination and its characteristics, and the demand is reflected by an origin, which considers visitors' attributes and the country or region where they come from (Uysal 1998). In the following sections, we look at both parts of the tourist system more closely and try to show how these concepts are connected.

2.1 Tourism system

2.1.1 Demand

According to Formica (2000), demand for tourism used to be examined by local policymakers as the only important part of the tourist system for a long time. They used it to explain, measure, and predict the economic benefits of tourism, such as revenues and employment, which stayed in particular at the center of their attention. Sinclair et al. (2003) stated a similar remark a few years later. They argued that less attention dedicated to tourism supply than demand was caused mainly by limited data availability and the nature of tourism product (not a single object, but a product composed of many goods and services provided). On the other hand, hundreds of papers were written on tourism demand topics already in the late 20th century. (Lim (1997) reviewed 100 studies on international tourism demand until 1994, Li et al. (2005) summed up to 420 papers considering tourism demand modeling and forecasting since the 1960s, Song & Li (2008) counted 119 different papers of this type only between 2000 and 2006)

Researchers decide on specific methods and explanatory variables based on research questions and according to their understanding of phenomena. Therefore, it is crucial to start by defining what demand in tourism precisely means. Song et al. (2010) provide a helpful and general definition of tourism demand. Principally, similarly to the ordinary concept of demand, it reflects the willingness to consume goods and services and the ability to afford them. However, tourism has some unique characteristics.

The first important thing is that consumers must travel to the point of supply if they want to consume tourism products and not the other way. Many researchers studying tourism demand revealed this point repeatedly (O'Hagan & Harrison 1984, Morley 1992, Song et al. 2010). One logical implication of this is that tourist commodity consumption could depend, among other factors, also on distance and accessibility from visitors' origins to a goal destination. In other words, for modeling purposes, both tourists' home and destination regions matter, which represents a typical tourist demand study problem. For that reason, authors sometimes focus only on travel between two or more countries (Lim & McAleer 2002, Malaysian demand for tourism in Australia; Dritsakis 2004, demand of tourists from Great Britain, Germany for holidays in Greece; Kusni et al. 2013, demand for Malaysia from OECD countries).

Next, a tourism product consists of complementary goods and services offered to the visitor as a whole package. Tourists request this as an entire bundle (O'Hagan & Harrison 1984, Morley 1992, Song et al. 2010, etc.). In other words, a tourism product incorporates various things created across different industries that build it all together. Which characteristics of a particular destination influence the tourism product significantly is examined by a tourism supply analysis.

As we wrote before, there are different possibilities for measuring and representing demand. Uysal (1998) claims that some of the most common variables are:

- Number of tourist arrivals or number of participants
- Amount of tourist expenditure and revenues
- Number of tourist's nights spent at the destination and their length of stay
- Travel propensity indices

Similarly, as no universal variable representing a volume of demand exists, we can also find various factors that could explain its value. Everything depends on the hypothesis one wants to test. Many researchers offer different ways to divide those variables into categories. Using Uysal (1998) explanation of tourism demand, he identifies three main categories: *Exogenous (business environment) determinants*, like business trends, economic growth, political stability and availability of supply resources; *Socio-psychological determinants*, such as attitudes about destination and factors, which affect the tourists' decision-making process; and *Economic determinants*, like income, prices and other typical economic factors.

Despite a large variety of potential variables, some are repeatedly selected in most studies. According to Lim's review 1997, 84 out of 100 papers about international tourism demand included variable *Income in the origin country*, which made it the most used factor. Other regularly used determinants were *Relative prices and tourism prices*, *Transportation costs*, or *Exchange rates*. The choice of this type of variables can be understood according to Uysal (1998) in fact that classical economics presents demand as a function of the price of goods and services, consumers' income and preferences, and prices of substitute and complementary goods.

The tourism product is by its nature heterogeneous. Destinations use various marketing strategies to differentiate from their competitors to show potential visitors they are unique and worth visiting. The above-mentioned typically used variables might be a good choice to start with in many cases. However, using only very general variables might not be enough. Also, important determinants of demand in one destination might not work in the other. The precise model of demand for tourism would be, without a doubt, very complex and might involve a vast number of variables. Therefore, some researchers decided not to analyze the tourism demand as a whole and instead directed their attention to certain types of variables.

An example is Cho (2010), who focused mainly on the non-economic determinants. Other studies are more specific and look for the causal effect of one phenomenon on demand (Moore 2010, the effect of climate change; Bassil 2014, the effect of terrorism; Balli et al. 2016, the effect of immigrants on demand for immigrant-source countries). This approach aims not to explain all the complexity of a system. Instead, it concentrates more profoundly on the contribution of its particular characteristics. We also decided on this setting in this thesis, specifically, the effect of cultural and natural heritage.

2.1.2 Supply

Studying tourism demand had a prominent position in the past because its results could be used in attempts to maximize the economic benefits of travel and tourism (Formica 2000). However, the booming mass tourism had developed some adverse aspects in some locations, in the form of so-called "overtourism." According to Koens et al. (2018), the difference between mass tourism and overtourism is that even though some locations can handle the increasing number of visitors relatively without problems, in other regions can also a slight change upwards bring significant negative influences. Formica (2000) sees the reason which changed the perception of tourism by policy-makers in the described phenomenon. In his view, because of the increasing problems due to mass tourism, this sector started to be evaluated regarding its integration with landscape and environment, which affected researchers' interest also to consider tourism supply in their works.

"Without tourist attractions, there would be no tourism", this borrowed idea from Gunn (1972) opens Lew's (1987) academic paper. Uysal (1998) argues similarly that the presence of tourism resources at destination sites is crucial for the existence of tourism alone. Formica (2000) claims that studies of tourism supply examine these existing tourism resources and the aim of these studies is usually to define how attractive a specific location objectively is.

Similarly, as in the demand analysis, supply factors can also be chosen according to multiple keys and divided into various categories. One useful academic paper, which covers some essential attributes of tourism and proposes a general categorization of tourism products, is written by Jafari (1974). Although this work is almost 50 years old, many new papers, also cited in this thesis, refer to this study (or the ideas presented by Jafari in some other his papers) as primary literature (Uysal 1998, Formica 2000, Kozak et al. 2008, McKercher 2016). They do so when defining what exactly can be meant under tourism supply. In the following part, we will hold on to Jafari's terminology, even though researchers sometimes use different terms. Therefore, we will call not the bundle of goods and services together a *tourism product*, but all the specific parts of it.

In a very general way, tourism products can be divided into *Tourism Oriented Products* (TOPs) and *Resident Oriented Products* (ROPs). As the name indicates, TOPs involve all economic activities focused on tourists and non-residents. These can be further divided into categories according to sectors, which they represent, specifically *Accommodations*, *Food Service*, *Transportation*, *Travel Agencies*, and *Other Travel Trade Services.* On the other hand, products primarily dedicated to residents but can also be used by tourists are ROPs. These could include most additional economic and other activities, such as *Health Services*, *Infrastructure*, etc. (Jafari 1974)

Tourists are not only attracted by goods and services offered to them but also and often primarily by various elements, which "create tourismagnetic atmosphere" (Jafari 1974, p. 77). These are called *Background Tourism Elements* (BTEs) and represent attractions of a destination. They can also be further divided into three different groups. *Natural BTEs* is the first one, and it includes all the natural resources and attractions, such as the sea, mountains, weather conditions, or beautiful natural sceneries. Another category is *Man-made BTEs*, including historical buildings and monuments, modern architecture, and many others. The last section is composed of so-called *Socio-cultural BTEs*, which are those elements which create a unique atmosphere and include some activities. There belongs, for example, culture and religion, celebrations and festivals, politics and so on. Similarly, as stated above, Jafari also concludes that these BTEs or attractions are the foundation of tourism as an industry and are the reason tourists travel to destinations. (Jafari 1974)

Among other works, which focus on the supply side of tourism and offer some categorization of its factors, we also mention Smith (1988). According to Smith, tourist goods and services can be divided into six groups: *accommodation, transportation, travel services, food services, activities and attractions, including recreation, culture, or entertainment*, and *retail goods.* These can be further classified into two tiers, which, similarly to Jafari's analysis, are composed of products oriented purely toward tourists and products with mixed usage.

This thesis will focus mainly on the products which play a crucial role in attracting tourists. The fact that a characteristic of "attractions," as was defined by Jafari (1974) , should be "to attract" visitors, can be derived already from its name. As was said above, opinions of experts could differ which all aspects of possible factors explaining tourism attractiveness can be marked as "attractions" and which have only a supportive role. Crouch & Ritchie (1999) argue that, for example, food services and accommodation could be considered as these non-prior reasons for travel. However, they emphasize that their quality can also significantly affect a tourist target's image.

A very intuitive implication that attractions should attract was challenged in the study by McKercher (2016). McKercher developed a framework for how the importance of attractions could be evaluated concerning their impact on tourism demand, specifically on the tourists' decision-making process. However, his work was only theoretical and was not applied to any real-world case. The question of whether what we consider to be an attraction plays a significant role in engaging people to visit some place is also central in this thesis.

2.1.3 Spatial properties of tourism supply

One attribute of the tourism system, which might be important to consider in analysis, is its interconnection with a specific area. Smith (1987) emphasizes the importance of geographical aspects of tourism and argues that they should have an essential role in researching tourism products. He developed a method to study tourism destinations on a regional level based on the types of tourism products they provide and how those products affect certain tourism demand measures. The analysis approach developed by

Smith was applied repeatedly in many newer papers with some variations and could be considered relevant primary literature to be mentioned in this thesis. Kozak et al. (2008) build on this approach and remember other works also doing so: Lovingood Jr & Mitchell (1989); Backman et al. (1991); Cha & Uysal (1995); Spotts (1997); Shoval & Raveh (2004). Many of these works bring a new perspective to the problem, state new questions, and adapt the model's attributes for a particular country or regional characteristics.

Depending on model settings, an analysis of spatial characteristics of destinations might not only be an exciting enhancement of a model but not including it might result in biased estimates. About consequences of not including analysis of spatial spillovers and heterogeneity in the models of tourism system (underspecified models) warn, among others, Yang & Wong (2012) and Yang & Fik (2014). The methodological part includes more information about different types of spatial effects and their mechanisms.

2.2 Attractiveness

2.2.1 Definition

As stated in the name of this thesis, the concept of tourism attractiveness is crucial for our analysis. One definition of this phenomenon comes from Mayo et al. (1981) and was further used among others by Vengesayi et al. (2009): "[destination attractiveness is defined as] the relative importance of individual benefits and the perceived ability of the destination to deliver these individual benefits" (Vengesayi et al. 2009, p. 622). On the other hand, Vengesayi et al. claim that it should be generally assumed that a destination, built as a combination of different tourist attributes, should become more attractive for tourists when more such attributes are added to it. From these thoughts, one can assume that the abstract term of attractiveness could reflect both the location characteristics and how this place is perceived.

One study trying to explain and create a model for the measurement of tourism attractiveness was written by Formica (2000). The author argues that two main methods for evaluating destination attractiveness used traditionally were analyzing it from a demand and supply perspective. A representative work of demand-side should be according to Formica (2000), for example, the one from Hu & Ritchie (1993), where it is assumed, similarly as by Mayo et al. (1981), that "the attractiveness of a travel destination reflects the feelings, beliefs, and opinions that an individual has about a destination's perceived ability to provide satisfaction in relation to his or her special vacation needs" (Hu & Ritchie 1993, p. 25). Formica (2000) believes that the supply method can be best described by Kaur's 1981 understanding of destination attractiveness, who should consider it as some pulling power based on the entire set of attractions located at a particular place at a specific time (Formica 2000).

Formica (2000) saw a gap in the availability of works studying tourism attractiveness from a general perspective when writing his analysis. He also tried to explain relationships between possible supply and demand indicators and further analyzed Nyberg's 1995 procedure. Nyberg's opinion was that it is appropriate to include the entire tourism system in analyzing attractiveness, specifically tourists, destination, and the linkage between these both supply and demand factors (Formica 2000).

2.2.2 Related concepts and problems in terminology

In addition to tourism attractiveness, other commonly used terms occur in related academic literature, and some of them might be ambiguous. In the following paragraphs, we will explain what will be precisely understood as attractiveness in this thesis and what will not.

One of such collocations is *tourism competitiveness.* Vengesayi's (2003) definition of tourism attractiveness originates in the before explained understanding of attractiveness by Mayo et al. (1981), looking at this phenomenon from the perspective of visitor's perception (demand-side). For the supply point of view, he uses the term *tourism competitiveness.* Vengesayi's understanding of competitiveness originates in Buhalis (2001), who uses similar terminology when analyzing tourism in Greece. The World Economic Forum also uses the expression *tourism competitiveness* in its reports since its experts regularly publish the so-called *Travel & Tourism Competitiveness Index*, which comprises 90 various indicators across 140 economies (Calderwood & Soshkin 2019). In this thesis, we will use the terminology by Formica (2000) and use the word attractiveness for both supply and demand perspectives.

Another expression, which is sometimes used together with tourism attractiveness, is *tourism potential*. Iatu & Bulai (2011) studied this concepts for the case of region Moldavia (Romania). They claim that this collocation arises in more significant amounts in the research papers after 1990, but its exact meaning and purpose have not been adequately defined. They also identified two main directions, how the tourism potential used to be understood. Iatu & Bulai (2011) call the first approach *immaterial* because the tourist potential is characterised as preliminary and describes what could happen at a specific location. The *material* approach describes destinations' actual situation and sums all the available tourist resources together. A similarity between commonly used expressions *attractiveness* and *potential* is also seen by Yan et al. (2017), who developed a mathematical model for measuring heritage sites' tourism potential. They argue that they prefer in their work the term *potential* over the *attractiveness* because the first one should indicate the state of a destination at the initial phase of tourism development, as they do in their work, and the other at any phase or on any scale.

The last concept, which gained rising attention, especially in the last years, is *tourism performance.* Assaf & Tsionas (2019) provided a review of the literature on this concept. They argue that the traditional simple measures of tourism performance, based on, among others, the number of tourism arrivals, were insufficient, as they should provide information only about some part of the tourism performance. The approach they defend could be characterized by focusing on the methods applying Data Envelopment Analysis (DEA) and the Stochastic Frontier (SF) to measure the efficiency of tourism destinations. Studies of this type are often regionally oriented and were used for analyzing many different locations around the world, for example: Barros & Dieke (2008) with hotels of Angola; Shang et al. (2010) with Taiwanese hotels; Barros et al. (2011) at French provinces level; Benito et al. (2014) with Spanish regions; Chaabouni (2019) for Chinese provinces; etc. Some works of this type also ask specific questions related to some concrete effects. An example is a work by Cuccia et al. (2016), which tries to find out using DEA what the effects of the inscription on the UNESCO World Heritage List are on tourism destinations' performance in Italian regions.

2.2.3 Methods for measuring attractiveness

As was stated in the last sections, many authors in the last years preferred analyzing the tourism attractiveness from both supply and demand sides simultaneously. Among those authors were also Formica and Uysal, who proposed a framework study in 2006. They claim in their review of previous research that "literature suggests that demand and supply independently or collectively may be used to measure tourism attractiveness" (Formica & Uysal 2006, p. 419). They base their work on the assumptions originating in the previous studies that the factors of both sides of the tourism system affect the tourism product "collectively and simultaneously" and together create the tourist experience (Formica & Uysal 2006, p. 419).

Their proposed method comprises nine steps leading to the final results. The supply side is evaluated objectively based on available data on attraction variables on observed locations, and it ends up with standardized attraction-dimension scores. The demand perspective is based on tourism experts' opinions on the related area, whose task was to distinguish the importance of specific factors of attractiveness and estimate the availability of attractions in the observed regions. The final attractiveness was evaluated concerning both two methods. The authors supposed this method can be helpful not only theoretically but also as a decision-making tool for tourism planners, marketers, etc. (Formica & Uysal 2006)

However, the whole method is relatively complicated because it uses different statistical methods and experts' opinions. Nevertheless, some of its parts could be effectively used in this work, having in mind that its goal is not to provide an exact model of tourist attractiveness but rather to evaluate the effect of certain attractions. Even though the concept and specific process used by Formica and Uysal are probably new, their main ideas are based on the supply and demand perspectives and approaches, which were known many years before. The authors categorized their used techniques based on the groups of methods described by Lew (1987).

Lew (1987) identified three different perspectives on studying tourism attractiveness. The first is called *Idiographic perspective* and is based on describing a destination's concrete attributes. These works often list all the relevant attractions in specific locations. It is chiefly used when analyzing small geographical units, such as cities. Tourism guidebooks are often used for identifying important attractions. It is also regular that authors of studies intend not to examine all the possible categories of attractions in the same deep but instead focus only on some of them. The second category is *Organisational perspective*. The main difference between this and the recent approach is that the organizational method examines the nature and certain general aspects and qualities of attractions rather than just their enumeration. *Cognitive perspective* is the last main one. While the Ideographic perspective was similar to the concept of tourism supply because it analyzes a destination's characteristics, this perspective focuses on tourists and their perceptions and experiences. There are also so-called *Cross-perspective measures*, which combine multiple perspectives. The different perspectives can build together a complex image of a particular location's attractiveness. (Lew 1987)

2.3 Role of cultural and natural heritage

Whether places with some culturally or naturally impressive sites should have an attracting power for tourists can be almost surely answered without long discussions. These factors were mentioned as one of the main attractors in the previous literature review, and their positive effect on attractiveness seems logical. Despite this, the relationships might not look that obvious when we look at the tourism reality in Slovakia. Some regions with relatively the highest amounts of historical and natural monuments belong to places with the lowest numbers of visitors per capita (more about those districts is written in the following sections Data and Results). Until now, we have discussed the effects of heritage sites only in theory. In the following, we summarise what the empirical literature tells us about their impacts.

Romão (2015) analyses the effects of culture and nature on tourism demand on a larger European scale (NUTS-2 regions). As proxy variables, the author uses the percentage of land registered in NATURA 2000 for natural sites and the number of locations inscribed on the UNESCO World Heritage List for cultural sites. Romão (2015) finds positive effects of both variables and also emphasizes the role of spatial interaction effects, which should be clearly recognizable mainly in Southern Europe. As a possible improvement, the researcher proposes using a smaller regional scale.

According to a study by Farid (2015), which examines two African states, there is a positive correlation between the existence of the World Heritage sites and the number of tourists. Another paper is written by Cuccia et al. (2016) and studies UNESCO World Heritage sites in Italy and the effect of the World List inscription on tourist performance (not attractiveness). Although the results show that cultural and environmental heritage has a positive effect on this performance, UNESCO's impact should negatively affect their efficiency (Cuccia et al. 2016).

In May 2014, the Council of the EU discussed the importance of cultural heritage at their meeting. As it stays in its report's name, cultural heritage should be considered "a strategic resource for a sustainable Europe" (Council of the European Union 2014). The role of cultural heritage is emphasized in the report because of various aspects, not only tourism, such as building and improving social capital, enhancing regional development, and creating jobs across different industries. This type of heritage should also give society many non-measurable benefits and externalities, such as enhanced quality of life, building identity, and increasing social cohesion (European Commission and Directorate-General for Research and Innovation 2015). The year 2018 was even declared as the European Year of Cultural Heritage. Cerisola (2019) went in her research even further and argued that the cultural heritage could inspire people to higher artistic and scientific creativity, which could have an indirect positive effect on economic development.

A concept of sustainable development and a linked term of sustainable tourism is undeniably related also to the natural heritage and its preservation. In their work, de Castro et al. (2015) examined determinants of the attractiveness of national parks in Brazil. Similarly, as we did, they decided to use the number of visitors as a dependent variable. They studied multiple potential independent variables, such as the reputation of a national park (based on the number of citations), diversity of activities, age of a national park, et cetera. Some other works, like Romao et al. (2013), are more concerned with the areas of protected lands, in this case, represented as a percentage of the territory inscribed in the European network Natura 2000. Other similar works involve the level of biodiversity (Siikamäki et al. 2015), or they use explanatory factor analysis to explain the effect of national parks, wildlife, and beautiful natural sceneries together (Vengesayi et al. 2009). The benefits of natural site preservation, especially of the Natural World Heritage, are also discussed in another report written for the IUCN (2014). According to this paper, the protection of the ecosystem and the World Heritage sites has many direct and indirect benefits for local communities. These positive effects should often be "associated with providing health and recreation values, knowledge, contributing to the local economy, and cultural and spiritual values" (IUCN 2014, p. 18).

Even though we are not aware that there are any studies analyzing specifically the effects of cultural and natural heritage on tourism attractiveness and demand in Slovakia, we found a few academic papers on related topics. Pompurova & Simockova (2014) analyzed the attractiveness of Slovak administrative regions from the point of view of tourists from the Visegrad 4 group of states. They found a great potential for visitations to grow.

Matijová et al. (2019) examined the relationships between different tourism indicators, such as the number of visitors or accommodation prices, and the unemployment rate.

3 Methodology

In this section, our goal is to explain the methodology behind the models used in this thesis. We will try to examine the logic of how the specific methods work and how we should interpret them. They will be set into the context of possible technics discussed in the last section, and we will try to clarify why we find these approaches to be valid and appropriate for our problems.

3.1 Towards a methodology

The main objective of our analysis is to estimate the effects of cultural and natural heritage on tourism attractiveness. In our analysis, we follow the understanding of tourism attractiveness as a phenomenon linked to both tourism demand and supply. For our purpose, we see no reason to analyze supply and demand separately and agree with Formica & Uysal (2006), that both sides of the system form the tourism experience together. The natural and cultural heritage is a typical example of an attraction and thus is a part of the tourist supply. The pulling power of attractions on a visitor (attractiveness) is represented by the number of accommodated tourists, a common variable representing tourism demand.

In the previous literature review, we already mentioned that these attractions might bring many possible benefits explicitly and implicitly, and tourists may admire them for various reasons. Therefore, a logical assumption is that the relationship between the presence of such attractions and tourists' willingness to visit places where they are located is positive. However, high positive correlation coefficient does not necessarily mean that the historical and natural monuments are the main reason for tourists to decide on a specific location.

As we mentioned already before, destinations with an abundant amount of attractive sites are usually all but not homogeneous. So are the tastes of tourism consumers. Many different factors, such as geography, infrastructure, destination awareness, and marketing, differ one location from another. All these unique characteristics make the study of tourism attractiveness, without a doubt, extremely complex. With a high number of different variables affecting the dependent variable, we might also assume collinearity between some of them, and by not including them, we risk the omitted variable problem. At the same time, we have a very limited set of observations, and therefore we cannot afford to include all potential variables in the model (not talking about data availability). For a valid estimation of our variables of interest, we must deal with these problems simultaneously. In order to do so, we must state one additional assumption.

We suppose that a distance between a destination and a visitor's home location will not significantly impact the attracting power of the place one wants to visit. The Slovak republic is relatively small in area. One can travel within a few hours to almost every attractive location. Important airports, such as Prague, Vienna, Krakow, or Budapest, make travel distance from the airport to selected tourist places relatively short. The infrastructure, such as railways and public buses, is also available. The fact that it is not possible to get information about tourists' origin on this regional scale of analysis, should therefore not affect results significantly.

It is generally known that a choice of correct independent variables may substantially impact the model and its explaining power. The interpretation of a model could also change after applying different variables to represent a similar phenomenon. For example, the surface of a protected area and the level of biodiversity of a natural park are linked to the same phenomenon, but the first describes a quantity and the second rather a quality of an area. The significance of estimates could also dramatically change. In reality, it might be sometimes impossible, or at least not very easy, to receive both qualitative and quantitative attributes of a phenomenon, especially when the thing is so abstract as the attractiveness is. We tried to keep it in mind when looking for a useful variable.

As mentioned in the literature review, the model of tourism supply and demand is by its nature linked to the spatial dimension, and not considering this in the analysis might potentially lead to biased estimates because attractions and tourism in one region might affect the neighboring regions. Elhorst (2014) argues in his book about spatial econometrics that there are usually two ways to start a spatial analysis. Most usual should be the one that starts with a standard non-spatial linear model, and the spatial interactions in the model are later tested and eventually included. The other one starts, on the other hand, with a more comprehensive model. We decided to follow the usual procedure.

3.2 Ordinary least squares models

A commonly used method for estimating linear regression is Ordinary Least Squares (OLS). A very general logic behind this concept is that the sum of squared residuals is minimized. An essential benefit of this model is that when all the necessary assumptions hold, the OLS estimator is "BLUE" - the best linear unbiased estimator. The type of data we are analyzing in the thesis is called cross-sectional, and since we are looking at the effects of multiple independent variables simultaneously, the model is called the multiple linear regression model. (Wooldridge 2013)

According to Wooldridge (2013), four specific assumptions must be met to get an unbiased estimator. The first one is called linearity in parameters, and as the name indicates, the model must have a linear form. Secondly, it is the random sampling assumption, which is, in our case, always valid, since we are operating with a population model (we do not need to sample our data because we are operating with relatively small datasets). Another assumption states that there may exist no perfect collinearity among the independent variables, which means that no regressor can be constant, and no exact linear relationship may exist between such variables. Although this assumption assumes only the perfect collinearity, very high collinearity can also be problematic, even though the estimators are not biased. Lastly, the zero conditional mean assumption must hold:

$$E(u \mid x_1, x_2, ..., x_k) = 0. \quad (1)$$

If this assumption is not met, it can mean that we have omitted some significant variables or we have some other misspecifications in the model. (Wooldridge 2013)

The following assumption is crucial to have the unbiased variance of the estimator. Under the homoscedasticity assumption, errors have the same variance irrespective of the value of regressors. (Wooldridge 2013)

The final assumption for proper testing of hypotheses using t and F statistics is called the normality assumption. It precisely says that "the population error u is independent of the explanatory variables $x_1, x_2, \ldots, x_k$ and is normally distributed with zero mean and variance σ^2" (Wooldridge 2013, p. 158).

As indicated, in our base model, we use the OLS method. We include only the most important variables there, representing tourism demand as dependent and attractions as independent variables. We suppose that the presence of both natural and cultural heritage should be positively correlated with the tourism demand. However, additional steps must be taken to determine whether the positive correlation can also be interpreted causally in the case of Slovakia.

We explained above that in order to get unbiased estimates, the first thing to check is whether we not omit any crucial variables. From the formula of the zero conditional mean, we know that we are looking for variables correlated to both dependent and independent variables, which omitting causes the model to over- or underestimate the effect of independent variables, and therefore creating the correlation between u and x to be non-zero.

By the nature of a problem, no test can discover the existence of omitted variables in a model (except functional misspecification, which is covered later). This model failure can usually be overcome by applying the theory of a studied phenomenon and potentially including additional variables until no omitted variables are longer expected. We continue accordingly and create a second, enlarged OLS model. Next, we perform several tests to check if all the assumptions are met.

3.3 Testing of Assumptions

3.3.1 RESET

Above, we mentioned a test that can reveal a specific form of a violation of the zero conditional mean assumption. Generally, this assumption can sometimes be the most problematic one. Wooldridge (2013) states that it is impossible to be sure there is no correlation between the explanatory variables and the average value of the unobserved variables. However, model misspecification might not only be caused by omitting a new variable, but the model can also be functionally misspecified, which can be relatively successfully uncovered. The presence of this failure says that the relationship between the observed independent variables and the response variable was not specified correctly. It could happen, for example, when some nonlinear relationships, such as squared forms and interaction terms of variables, which exist in the dataset, were not included in the model. (Wooldridge 2013)

The test, which can reveal the general form of these problems, is called Regression Specification Error Test (RESET), and it is often called also after his founder James B. Ramsey (1969). The idea behind the model is that we include polynomials in the OLS fitted values to the basic model, and we use F statistic to test a hypothesis. We can choose how many functions of the fitted values we want to add into the extended model, but typically, we test the model up to the third-order polynomial. The formula of the extended model is:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^2 + error \tag{2}$$

The null hypothesis is, that $\delta_1 = 0$ and $\delta_2 = 0$. If the F statistic is significant, there probably might exist some functional form misspecification. (Wooldridge 2013)

In 2008, Baggio published a study that criticizes a traditional approach of many researchers to study mainly the linear relationships between variables of tourism systems. He argues that "tourism destinations behave as dynamic evolving complex systems" (Baggio 2008, p. 1), and emphasizes the interdependence and high level of non-linearity between them. We are aware of these facts, and we tried to overcome parts of this problem by applying some basic nonlinear functional forms of the data. However, since the analysis of complex systems is probably far beyond the scope of this bachelor thesis, we will remain by using the traditional linear approach, even though the more complex procedure would probably provide better results.

3.3.2 VIF

We emphasized six assumptions that need to be fulfilled for a model to be BLUE and have a valid inference. The first three should be fulfilled now because we construct the model to be linear in parameters, examine the population, and the zero conditional mean should hold.

The following assumption deals with perfect collinearity. One statistical tool used to find excessive multicollinearity is called Variance Inflation Factor (VIF). It assigns a specific value to all of the model's independent variables based on the correlation level between a particular variable and other regressors. This statistic can be derived from the formula of variance of an estimate. (Wooldridge 2013)

Wooldridge argues that there is no universal cut-off value, indicating that multicollinearity is a problem and that the standard deviation of an estimate is too high. In the literature, the value of 10 is considered a warning signal that a correlation between variables might be too high and cause a problem. The statistic has, however, a more informative character.

(Wooldridge 2013)

3.3.3 Tests for homoscedasticity and normality assumptions

The OLS estimates should now be unbiased. To say the same also for the variance of the estimates, we must add the assumption of homoscedasticity to our model. We can test this assumption using different tests, among which the Breusch-Pagan test for heteroscedasticity (BP-test) and White test for heteroskedasticity belong to commonly used and are described in more detail also by Wooldridge (2013). The difference between those two methods is that the White test is more general and can uncover various other forms of heteroscedasticity, which the BP-test cannot find. The White test includes additional functions and combinations of independent variables. Those functions, together with all independent variables, are estimated in a regression, where a dependent variable is built as squared OLS residuals. We use the F or LM statistic to test the null hypothesis that coefficients of the included regressors are equal to 0, which implies homoscedasticity. (Wooldridge 2013)

If all the previous assumptions are met, we can move to the last one, normality. Wooldridge (2013) denoted this assumption as stronger than the previous because it assumes the zero conditional mean and homoskedasticity assumptions. All the six assumptions together are called *classical linear model (CLM)* assumptions, and if they are fulfilled in the case of cross-sectional data, we can talk about valid inference. Wooldridge (2013) also explains that finding out whether errors are normally distributed is usually an empirical problem. In practice, we might try to state whether the distribution of errors is at least "close" to normal (Wooldridge 2013, p. 120). For that reason, we will prove this last assumption mainly empirically by looking at the residuals graph and comparing it with the normal distribution graph.

3.4 Spatial interaction effects

3.4.1 General spatial model

At this point, the usual procedure of preparing the model based on linear regression would be completed. However, in our case, the observations are created by spatial units. This fact adds some additional specifications to be considered by the analysis, which we will discuss now.

Kelejian & Piras (2017) state at the beginning of their book Spatial Econometrics that the spatial models are generally based on the intuitive principle that the bigger the distance between some observations, the less connected they are. The observations have their spatial context, and closer observations might affect each other. Its implication for our model is that demand for tourism in one location might be affected by attractions and demand in another location. The proximity of specific observations can be described in the so-called *neighborhood weighting matrix*. There are different ways how this closeness can be specified. In this thesis, we will discuss two of them. The first type we will use depicts only information on whether two regions are adjacent. Another type provides information about the distance between two observations. (Kelejian & Piras 2017)

Choosing a type of weighting matrix would probably depend much on the type of spatial information we want to uncover. Elhorst (2014) summarizes in his book some most essential knowledge about this field of econometrics. He explains that there are three main types of spatial interaction effects for cross-sectional data:

The first interaction effect could occur when values of response variables are interdependent across space. These interactions are called endogenous. The second type is called exogenous, and as the name refers, the regressand values are, in this case, dependent on the values of regressors for nearby regions. Lastly, the interaction effects can operate between different observations' error terms. (Elhorst 2014)

There is no rule that these effects cannot be presented in a model simultaneously. The classification of the linear spatial dependence models is derived from the general model, which assumes all three effects. Elhorst (2014) defines the general formula as follows:

$$Y = \delta WY + \alpha + X\beta + WX\theta + u \qquad u = \lambda Wu + \varepsilon \tag{3}$$

where the variables Y, X represent the dependent and independent variables respectively, and u is a disturbance. ϵ is an error term, defined as usually. Since we work with matrices, all the variables should be understood in terms of matrices. W is a weighting matrix of type $N \times N$, where N is the number of observations. The intercept expression is represented as $\alpha \times 1$ vector of length N. WY represents the endogenous interaction, WX the exogenous, and Wu the interaction among disturbances. The coefficients next to these variables are also defined in matrix form according to their interactions. (Elhorst 2014)

We can see in the model that if the coefficients $\delta = \lambda = \theta = 0$, we obtain the usual linear regression. We can also assume that just some of these coefficients are equal to 0 and obtain other specific types of spatial models. We start with the case, which does not differ from the usual OLS in the assumptions.

3.4.2 SLX

The model is called the Spatial Lag of X (SLX) model, and it assumes only the exogenous spatial interaction effects. When we look at the above-mentioned general model, we suppose that $\delta = \lambda = 0$. Its formula for a model with one explanatory variable is: (Elhorst 2014)

$$Y = \alpha + X\beta + WX\theta + \varepsilon \tag{4}$$

According to Halleck Vega & Elhorst (2015), this model has been often ignored by theorists of spatial econometrics compared to other types of spatial models, even though it has been used repeatedly in applied research papers. The authors advocate this method and its usage, and they recommend it to select as a "point of departure" model, especially when there are not enough solid theoretical arguments supporting the choice of another specific spatial model (Halleck Vega & Elhorst 2015, p. 360). They further explain that econometricians do not often focus on the SLX model because it can be estimated using standard procedures known from the OLS model.

3.4.3 SAR

Spatial autoregressive model (SAR), also called the spatial lag model, is another spatial econometric method. Similar to the SLX, also this model involves just one type of interaction effect. We call it an endogenous interaction effect because values of the dependent variable of neighboring regions affect each other (spatial lag). The formula is following: (Elhorst 2014)

$$Y = \delta WY + \alpha 1_N + X\beta + \varepsilon \tag{5}$$

In comparison to the SLX, this model is more complicated for the interpretation of spatial effects. Generally, Elhorst (2014) identifies two types of spatial effects common for all types of spatial models. To show them, we first rewrite the general nesting spatial model, which includes all three types of spatial interactions:

$$Y = (I - \delta W)^{(-1)}(X\beta + WX\theta) + R \tag{6}$$

where R involves the other variables of the general model, not important for the interpretation (Elhorst 2014).

Following Elhorst's (2014) explanation, we want to find the expected value of Y to define direct and indirect effects. We build a matrix of its partial derivatives with respect to the kth explanatory variable for all N observations:

$$\begin{bmatrix} \frac{\partial E(y_1)}{\partial x_{1k}} & . & \frac{\partial E(y_1)}{\partial x_{Nk}} \\ . & . & . \\ \frac{\partial E(y_N)}{\partial x_{1k}} & . & \frac{\partial E(y_N)}{\partial x_{Nk}} \end{bmatrix} = (I - \delta W)^{(-1)} \begin{bmatrix} \beta_k & w_{12}\theta_k & . & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & . & w_{2N}\theta_k \\ . & . & . & . \\ w_{N1}\theta_k & w_{N2}\theta_k & . & \beta_k \end{bmatrix} \quad (7)$$

where besides usual notation w_{ij} is the (i, j)th element of the weighting matrix W. The diagonal elements of the matrix above serve as a direct effect, and the other (off-diagonal) elements stand for an indirect effect. If no spatial interactions take place and the standard OLS model is applied, there are no indirect effects, and the direct effect is β_k. The same is also valid for the model when only spatial interactions between errors occur. SLX model has the same direct effect as the OLS, but it also involves an indirect effect represented by the θ_k. That is the reason why an SLX model does not differ much from a usual OLS model with additional independent variables. From the rewritten matrix of partial derivatives above, only the part inside the brackets remains in the equation because δ is equal to 0. The indirect effects of this θ are called *local spillover effects*. (Elhorst 2014)

The effects of the SAR model are, however, more problematic than the other. The rewritten equation with partial derivatives takes the following form:

$$\begin{bmatrix} \frac{\partial E(y_1)}{\partial x_{1k}} & . & \frac{\partial E(y_1)}{\partial x_{Nk}} \\ . & . & . \\ \frac{\partial E(y_N)}{\partial x_{1k}} & . & \frac{\partial E(y_N)}{\partial x_{Nk}} \end{bmatrix} = (I - \delta W)^{(-1)} \beta_k = (I + \delta W + \delta^2 W^2 + \delta^3 W^3 + ...) \beta_k \quad (8)$$

The diagonal elements of weighting matrices W are equal to zero, whereas the identity matrix I has the 0-values everywhere except diagonal. Therefore, the first two terms in the decomposed matrix on the right (β_k and $\delta \beta_k W$) represent the direct effect and indirect effect on first-order neighbors, respectively. However, the whole effect does not solely depend on these terms since there are also other terms of higher orders. As a result, also not directly connected or remote regions are reached. For that reason, the indirect effects of this type of model are called *global spillover effects*. Elhorst (2014)

3.4.4 SEM and SAC

We discussed spatial lag of independent and dependent variables, so the last type of spatial dependence is a spatial lag in error. We start again with the general formula, using the same notation as before. The basic model including only this spatial interaction is called the spatial error model:

$$Y = \alpha + X\beta + u$$
$$u = \lambda W u + \varepsilon \tag{9}$$

as showed by Elhorst (2014).

Spatial errors can be present if some factors influencing the response variable are omitted and autocorrelated through space or when we do not involve effects of a shock operating spatially to the model (Elhorst 2014). Bauhoff (2005) explains that this concept is comparable to serial correlation for time-series data, and it affects only efficiency, not the value of an estimate.

For some problems, we might be interested also in models which combine various types of spatial interactions. In our case, the method of spatially lagged independent variables might not be very useful because of a limited number of observations. Since the SLX model doubles the number of independent variables, we might obtain an overfitted model. Therefore, the only combined model worth mentioning for our purposes is the SAC model, linking SAR and SEM models. This model would have the same formula as the general model with all spatial effects except θ=0. Elhorst et al. (2013) argue that the characteristics of direct and spillover effects of this model are identical to SAC. For that reason, we do not go more into detail.

3.4.5 Spatial models estimation method

We mentioned that the SLX model could be estimated similarly to the standard linear regression using OLS. Chi & Zhu (2019) claim that an appropriate method for estimation of other spatial econometric models (SAR, SEM) is the maximum likelihood estimation (MLE). Elhorst (2014) also mentions other approaches besides MLE, such as quasi-maximum likelihood, generalized method of moments, and Bayesian Markov Chain Monte Carlo methods. In our models, we use the *lagsarlm* function from the *spatialreg* package in R language, which operates with the maximum likelihood estimation (Bivand 2022). Therefore, we introduce this estimation method briefly.

Wooldridge (2013) provides a short introduction to the MLE method in the appendix of his book. Instead of minimizing the sum of squared residuals of a model (as we do with OLS), this type of estimator maximizes its log-likelihood function. When done properly, this estimator is consistent and asymptotically efficient, and its variance is the smallest possible compared to all other unbiased estimators (OLS is a special case of MLE). Compared to the OLS, we do not need to assume a normal distribution. Nevertheless, we need to assume the distribution of the dependent variable, conditional on independent variables. This method also allows the analysis of more complicated econometrics models, such as those including non-linear relationships. (Wooldridge 2013)

3.4.6 Testing for spatial dependence

We introduced different ways, how estimates can be affected by spatial dimension. However, it is not a rule that spatial effects must be present in the regional data. Kelejian & Piras (2017) describe different tests serving to discover spatial correlation. Among them are also the *Lagrangian multiplier tests*, which are often used also in other than spatial types of econometrics. We decided to use this type of tests in our thesis, because of its universal principles known from traditional econometrics.

Kelejian & Piras (2017) provide also a general logic behind this test. It should build on likelihood maximization, and comparing the restricted and unrestricted model. A model should be restricted under the null hypothesis. In our case, in the general spatial model we test under the null hypothesis, whether $\delta = 0$ or $\lambda = 0$ (or both together). (Kelejian & Piras 2017)

3.4.7 Weighting matrix

As mentioned at the beginning of this subsection, the concept of a weighting matrix is crucial for spatial analysis because it defines the neighborhood of observations. If we have N observations, the weighting matrix, denoted by W, will have N rows and N columns corresponding to the observations in the same order. Let us call the element in the i-th row and j-th column w_{ij}. This element describes the closeness between the i-th and j-th observation; provided i and j were chosen from $1 \dots N$. The element in the i-th row and i-th columns, w_{ii}, will be equal to 0, for $i = 1 \dots N$, because we do not consider an observation to be a neighbor of itself. Kelejian & Piras (2017)

As mentioned before, the first type is the adjacency matrix. One could also encounter this type of matrices in other related study areas. The logic is elementary. The value 1 indicates that two units are neighboring, and 0 means the opposite. Kelejian & Piras (2017)

The second type is the inverse distance matrix. Kelejian & Piras (2017) state that many researchers use to transform the information about the distance in the following way:

$$w_{ij} = \frac{1}{d_{ij}} \tag{10}$$

where d_{ij} represents the distance between two specific observations. The reason is also straightforward since we usually want to interpret the weights so that more distant observations are assigned lower weights. The weights of very remote units should thus converge to 0. (Kelejian & Piras 2017)

Elhorst (2014) explains that researchers often use the row-normalized version of a weighting matrix (sum of all elements in every row is 1) because one can interpret the resulting values as an average effect of neighboring observations. Even though Elhorst also sees some drawbacks of this practice, we think it might be an appropriate method for our problem. The reason is the geography of Slovakia and its administrative division. Slovakia is a mountainous country, and its population is unequally dispersed. It causes some districts located mainly in the mountains in the central part, like Brezno, to be relatively large and have even ten direct neighbors, whereas others, like Skalica, lying at the border, have only two. At the same time, Slovakia is landlocked and is part of a Schengen control-free travel area. Therefore it can be easily accessed from most parts of the border. Not using row-normalized matrices would mean that districts like Brezno could have relatively larger values of lagged variables than districts like Skalica, even though we think it is improbable that regions at the border are not affected at all by tourism regions on the other side of the border. In fact, the district of Brezno is locked by mountains from many sides, so it might be harder to access.

4 Data

In the following pages, we present the observations and discuss the dependent and independent variables of the models. We explain how we proceeded in the data selection and describe some specifications of the dataset. We also provide some descriptive statistics of the variables.

4.1 Observations in general

We decided to quantify the effect of cultural and natural heritage on tourism in Slovakia on the data from 2018. We wanted to examine as new data as possible so that the outcomes might also be relevant these days. The years 2020 and 2021 were hardly hit by the pandemic, and especially the tourism sector belongs to the most affected industries. The year 2018 was, therefore, a natural choice because we could compare it also with the previous and following years, 2017 and 2019, and most variables did not change much, as will be shown.

As an observation unit, we chose the level of districts (*okresy*). Slovakia is administratively divided into 79 districts. The capital Bratislava consists of five different districts called Bratislava I – V, and in the same way, are also called four districts of the second largest city – Košice (Košice I – IV). The reason for this choice is that districts are the smallest possible unit for which we could find all the relevant variables. Using larger units would result in a lower number of observations, which would make our models less confident.

Our final dataset consists of 71 observations. Firstly, we summed up all the four districts of Košice to create just one observation representing the city. One reason is that it probably cannot be assumed that tourists visiting Košice for recreational purposes accommodated in, for example, the district Košice III are attracted primarily by attractions in this district and not by the center of the city. The other reason is that some statistics are provided only about the city as a whole.

Secondly, we decided not to include all Bratislava districts. The capital of Slovakia affected the analysis in some statistics as an outlier, which means that the presence of Bratislava in the data would distort the effects of individual variables. It is also probable that some other variables influence visits to the capital, which are not present in such amounts in other cities, such as business and shopping trips, visits to important institutions located there, good connections to the near metropolis of Vienna, better recognition by foreign visitors, et cetera. On the other hand, Bratislava, as an important center, might impact the neighboring regions, and not including those effects might also result in biased estimates. Therefore, we included a variable representing the distance from Bratislava to uncover them.

4.2 Dependent variable

In the literature review, we mentioned multiple ways to measure tourism demand. From the four most common variables, we managed to obtain two (the third presented variable is connected to the first two and can easily be derived) published on our local scale annually:

- **Number of tourist arrivals** - number of visitors in accommodation establishment of tourism in total: absolute numbers in levels for each district for year 2018 (Statistical Office of the Slovak Republic 2018*c*),
- **Number of overnight stays** - number of nights spent by visitors in accommodation establishment of tourism in total: absolute numbers in levels for each district for year 2018 (Statistical Office of the Slovak Republic 2018*c*),
- **Average number of overnight stays** – per one visitor, calculated as overnight stays divided by the number of arrivals.

Despite all of these variables might, in some context, represent the tourism demand, they describe different information. The average number of overnights is without a doubt necessary for tourism planners because longer visits might usually mean more money spent by a visitor in a destination. However, visits throughout the longer periods are probably often at those locations where visitors stay due to medical or therapeutic purposes, such as spas. Our data also support this claim: the Slovak average is 2.83; the districts with popular spa resorts are Bardejov (6.84), Turčianske Teplice (6.14), Piešťany (5.47); some of the generally known most popular tourist destinations in Slovakia are Liptovský Mikuláš (2.67), Košice (2.08), Banská Štiavnica (1.86). Therefore, our main objective to examine cultural and natural heritage effects would probably not be explained by this variable very well.

We decided to use the number of arrivals rather than the number of overnight stays in this thesis. We suppose that the second one would be a more relevant variable if we were looking for the performance of a tourist destination. However, the attractiveness might not necessarily assume a long time spent at a certain location. Some tourists prefer short holidays or weekend trips to attractive locations. Tourist agencies are often aware of this fact, and many metropolises offer visitors city cards valid for a few days. Even the official slogan of the Bratislava Tourist Board is "Bratislava - 72 hours city" (Capital City of Slovakia - Bratislava n.d.).

We found some missing observations for the dependent variable in the original dataset from the Statistical Office of the Slovak Republic. However, we found ways how these values can be found or calculated. We start with missing values for the districts Košice II and III. Since these were the only missing observations in the more extensive Košice Region (consisting of 11 districts), and the value for the whole region was available, we easily computed the number for Košice as a whole city by subtracting all non-Košice districts from the value for whole Košice Region.

The last two other districts with missing data were Humenné and Medzilaborce. Because both are located in the Prešov Region, we had to think of other solutions. Luckily for us, the ŠÚ SR published data for both districts in the following year, 2019, and the proceeding years, 2016 and before (2017 and 2018 were unavailable). Therefore, we could observe some trends to find out that values between 2016 and 2019 did not change much. We calculated the averages between these two years. Then, we subtracted the other districts' values from the value of a whole region. The difference represented the sum of the values for these two districts. The result was almost precisely the same as the sum of averages between 2016 and 2019 for the two missing districts. We divided the minor difference by two and added them to both obtained numbers. Thus, we ended up with no missing values.

4.3 Independent variables

4.3.1 Cultural and natural heritage

Starting with the cultural heritage, we could choose the variable depending on scale and quality. From the scale point of view, we can distinguish between protected areas and specific monuments. The protected areas can be further divided according to their value to those with the higher level of protection called *monument reservation (pamiatková rezervácia)* and lower called *monument zone (pamiatková zóna)*. The protected monuments have only one level of protection, and they are called *national cultural monuments (národná kultúrna pamiatka)* - NKP (there are also some other divisions and categories, which are not compatible with our focus). We prefer the more minor scale data in the analysis because there are hundreds of the NKPs in various districts, but only a small amount of the monument reservations and the monument zones (for example, no district has more than two monument reservations). We suppose that using the number of NKPs, we can better describe differences in the abundance of cultural heritage and so more precisely estimate the effect of their attractive power:

- **Number of National Cultural Monuments** (variable "Cult") - in absolute numbers for each district for year 2020 (we obtained only data for this year, but we know that the register of monuments changes only slightly throughout the years - can be considered as proxy for cultural monuments supply) (The Monuments Board of the Slovak Republic 2020)

A similar distinction can also be made in the case of natural heritage, except for a few specifications. According to the size of a protected area, the largest are highly protected *national parks (národné parky)* and less *protected landscape areas (chránené krajinné oblasti)*. As we mention, a common practice in the academic literature is to use the area of national parks as a variable representing natural attractions. However, tourists do not usually visit all parts of parks equally and are not allowed to walk outside from tourists' paths. Furthermore, natural parks are usually large in area and are located in multiple districts simultaneously, but the most visited places might be present only in some of them. Therefore, our opinion is that looking at a smaller scale would also be a more appropriate strategy in this case.

Another level involves such smaller areas to which belong *national nature reserve (národná prírodná rezervácia)* - NPR and *nature reserve*, and even smaller are *national nature monument (národná prírodná pamiatka)* - NPP and *nature monument*. The logic is the same as before: the qualifier "national" means that the area is the most valuable. We decided not to study the not-national units because there are many of them, and already some more valuable national monuments and reservations are often relatively unknown locations. The NPRs and NPPs can also be a part of larger units (national parks and protected landscape areas), so we will not lose the information about large areas. At the same time, large areas are usually divided into multiple neighboring NPRs and NPPs (such as in the case of the High Tatra). Since the natural waterfalls and caves belong to the monument categories and not the reserves, and they also often belong to popular tourist destinations, we think it might make more sense to quantify the effect of natural heritage as a sum of separate areas and monuments tourists can potentially visit. Our final explanatory variable would thus be:

- **Number of National Nature Reserves and National Natural Monuments** (variable Nature) - independent variable, described as a sum of the number of nature reserves and natural monuments summed over the districts for the year 2020 (a proxy for the supply of natural wonders for the same reason as the NKPs) (Štátna Ochrana Prírody Slovenskej Republiky 2020), An interesting variable representing a site's extraordinary quality and uniqueness could be its presence on the UNESCO World Heritage List. In Slovakia, there are several cultural monuments inscribed and also two natural areas. However, the problem with this variable is its heterogeneity and complicated administrative division. A good example is the district Banská Štiavnica which is inscribed on the list as a town together with the landscape in the vicinity. Small parts of it are also located in the neighboring districts. For a different reason, the Carpathian wooden churches are also in multiple districts. They do not cover almost any neighboring area but are scattered in various villages throughout Slovakia. As a result, we decided not to include this variable.

4.3.2 Other variables

It is necessary to have a good knowledge of the tourism situation in all districts in order to include all the relevant variables in the model and so obtain unbiased estimates for the main explanatory variables. It was also the main reason to study Slovakia because we are most familiar with this country. An essential source of information was also the website of the official Slovak travel agency: Slovakia Travel (n.d.), which presents highlights of each region. Very useful tools in looking for sources of potential endogeneity were residuals. Based on their structure, we could more easily discover omitted variables causing over- or underestimation. More about the whole process will be written in the following main section after presenting the results of the first model.

The additional variables are the following:

- **Attractiveness of Ski Resorts** (variable Ski) – obtained as a sum of indices ranging from 0 to 1 assigned to each ski resort over the district, taking into account their various aspects. Such a variable shall serve as a proxy and mirror ski resorts' quality, size, and attractiveness. Indices were taken from the analysis performed in Bučeková et al. (2019) (we used only resorts with an index of at least 0.2 to eliminate very low-quality ski slopes);

- **Sales of Medical Spa** (variable Spa) – should serve as a proxy to describe the quality and attractiveness of medical spa resorts. Sales are summed for each district. Data were collected for the year 2019 by TREND Analyses (2020). Value for the spa in Piešťany also included sales from the spa in Smrdáky, located in a different district. To find an approximate ratio of sales from the two spa resorts, we used their sales from 2014 published in Gúčik et al. (2016);

- **Population** (variable Pop) – the number of inhabitants of each district for the year 2018 (Statistical Office of the Slovak Republic 2018*b*);

- **Unemployment rate** (variable Unempl) – the unemployment rate in percents per district in 2018 (Statistical Office of the Slovak Republic 2018*a*);

- **Distance from Bratislava** (variable BA_Dist) – representing a distance between the main town of a particular district and Bratislava in kilometers (the distance matrix was adjusted by and used in Kotrč (2020), originally downloaded from Ľudmila Jánošíková (n.d.))

We created the spatial lag versions of the variables in R language using the above mentioned distance weighting matrix and the adjacency weighting matrix, both provided by Mgr. Michal Kotrč.

4.3.3 Descriptive statistics and adjustments

Table 1 shows the descriptive statistics of the variables as provided by the sources and described in the last sections.

Table 1: Unadjusted variables - descriptive stat.

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Visitors	60,652.8	95,903.0	430	15,817.5	28,895	70,965.5	612,181
Cult	130.4	115.4	17	53	93	164.5	566
Nature	4.2	5.6	0	1	2	5	26
Ski	0.1	0.2	0.0	0.0	0.0	0.0	1.3
Spa	2,070.2	5,308.1	0	0	0	0	32,109
Pop	70,669.8	43,158.9	11,896	37,937.5	62,286	95,255	238,757
Unempl	5.8	3.7	1.7	3.0	4.4	8.1	16.1
BA_Dist	227.1	132.7	21.8	117.0	205.2	335.0	475.9

After we chose our dependent and independent variables, and solved the problem with missing observations, we looked at the data more closely. Starting with the dependent variable, we found two issues which can be overcome by a transformation.

Firstly, we observed that size of districts determines number of visitors to the large extent. The more populous districts generally have more visitors. The reason is probably similar to the case of Bratislava - regional centers. Nevertheless, those districts are not such outliers as the Slovak capital and are a natural part of the dataset. We might solve this problem by including a variable *population* to the model as a proxy to all variables distinguishing large districts/cities from small (in cities, we might assume more overnight stays due to work trips, family celebrations - more people living there, nightlife, etc.). However, as we found in the model preparation phase, this relationship is quite logical and stays not in our focus but increases R-squared significantly. Even worse would be including *capacities of accommodation facilities*, which is also without a doubt an important determinant of tourism demand, but accommodations are usually not the main reason to travel, and the variable is extremely correlated with the number of visitors. Therefore, we decided to use the dependent variable normalized by the number of inhabitants: ***Visitors per capita * 1000*** *(VpC)*. The variable is multiplied by 1000 to be more readable.

Secondly, the values of the dependent variable are highly skewed to the right - only a small fraction of the districts have a relatively large number of visitors. Even though normalizing by population reduced this problem slightly, it is still evident, as can be seen in the Figure 1. This fact affects the fits of models, and some relationships are not linear or are too affected by large values. We found a better way to fit the data in logarithmic transforma-

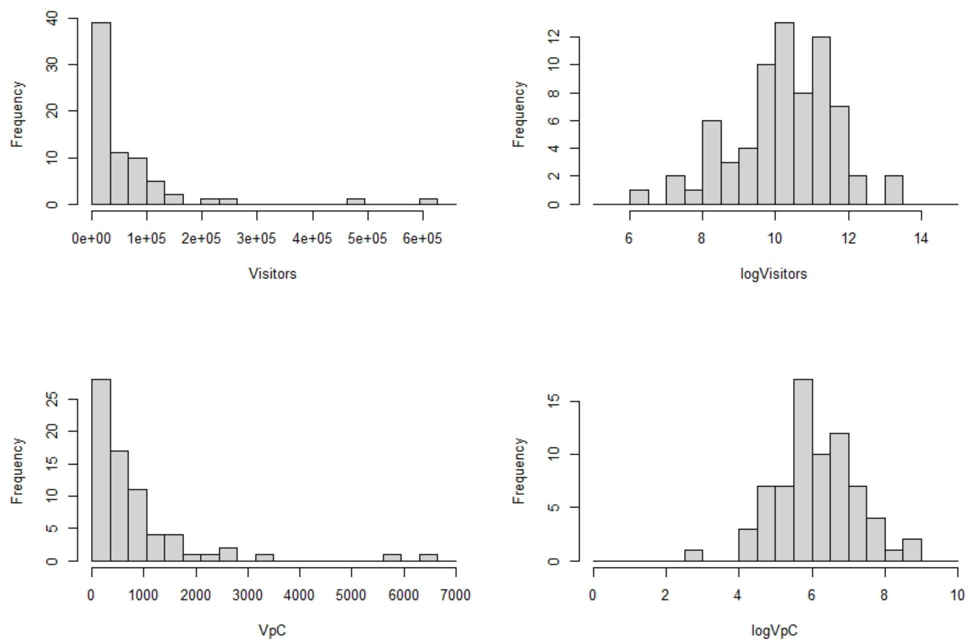

Figure 1: Histograms of the dependent variables

tion. We applied natural logarithm and created our final dependent variable: ***Logarithmic transformation of VpC*** *(log_ VpC)* (VpC is defined above).

As a next step, we looked at the relationships between the dependent and independent variables to determine whether the transformation made the relations more linear (Figure 2). An essential difference can be seen for both variables, Cult and Pop. In both cases, we could observe relatively lower variance for small values of VpC and Cult/Pop but higher for large values. After the transformation, the heterogeneity problem is no longer visible, but we can observe curves typical for logarithms - both variables are good candidates for the log-transformation. Another new final variables thus would

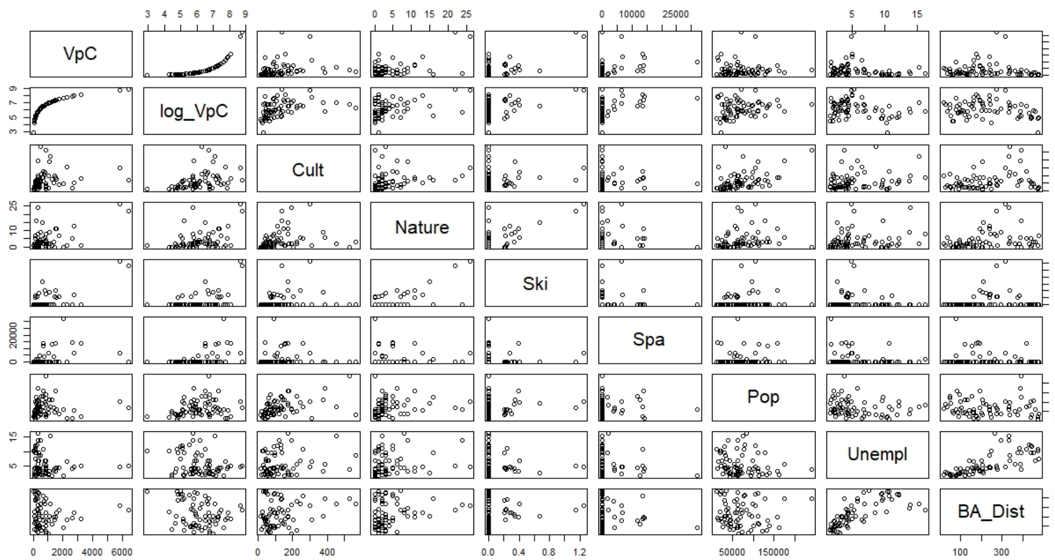

Figure 2: Relationships between pairs of variables - 1

be: ***Logarithmic transformation of Cult*** *(log_Cult)* and ***Logarithmic transformation of Pop*** *(log_Pop)*.

In addition, we look at the relationship between log_VpC and BA_Dist. Even though the linear relation might work in this case quite well, a potential improvement could be including the quadratic form of this variable. Especially interesting might be the interpretation of this because we could see that the variables near and far away from Bratislava generally have lower values. We will look at this nonlinear effect in the discussion. In one model, we also include the variable: ***Squared value of BA_Dist*** *(BA_Dist_2)*.

For the variables, Nature, Ski, and Spa is the most observations located around 0, and we do not clearly see the curve typical for logarithm in the data. Figure 3 shows the relationships with the new variables log_Cult, log_Pop, and BA_Dist_2. Table 2 involves the descriptive statistics of the variables used in the OLS models after adjustments. The descriptive statistics of the new spatial lag variables used in the second part of the analysis are shown in Tables 3 and 4. The values of observations are visualized on the maps in Appendix A.

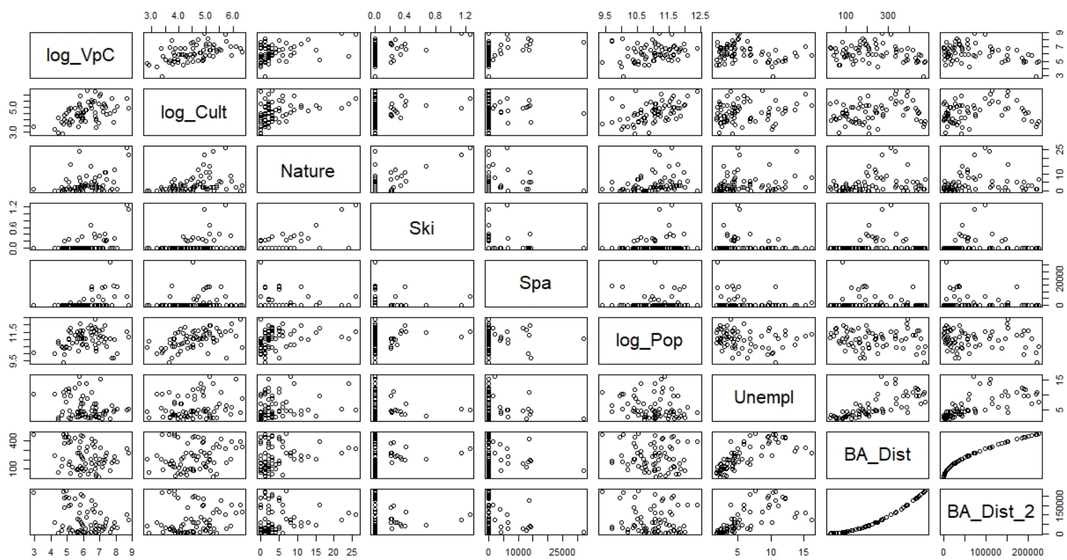

Figure 3: Relationships between pairs of variables - 2

Table 2: Adjusted variables - descriptive stat.

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
VpC	837.2	1,132.3	18.8	224.8	418.3	882.7	6,434.7
log_VpC	6.1	1.1	2.9	5.4	6.0	6.8	8.8
log_Cult	4.6	0.8	2.8	4.0	4.5	5.1	6.3
Nature	4.2	5.6	0	1	2	5	26
Ski	0.1	0.2	0.0	0.0	0.0	0.0	1.3
Spa	2,070.2	5,308.1	0	0	0	0	32,109
log_Pop	11.0	0.6	9.4	10.5	11.0	11.5	12.4
Unempl	5.8	3.7	1.7	3.0	4.4	8.1	16.1
BA_Dist	227.1	132.7	21.8	117.0	205.2	335.0	475.9
BA_Dist_2	68,925.6	68,641.3	475.2	13,726.2	42,107.0	112,267.8	226,480.8

Table 3: Spatial lags (adjacency matrix) - descriptive stat.

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
lag_log_Cult	4.6	0.5	3.6	4.3	4.6	4.9	5.8
lag_Nature	4.8	3.9	0.8	1.9	3.2	6.6	16.0
lag_Ski	0.1	0.2	0.0	0.0	0.0	0.1	0.6
lag_Spa	2,061.3	2,358.7	0.0	0.0	1,405.7	3,403.5	8,988.2
lag_log_Pop	11.1	0.3	10.5	10.9	11.1	11.3	11.8

Table 4: Spatial lags (inverse-distance matrix) - descriptive stat.

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
lag_log_Cult	4.6	0.4	4.0	4.3	4.5	4.7	5.6
lag_Nature	4.1	2.5	1.4	1.9	3.2	5.3	10.8
lag_Ski	0.1	0.1	0.0	0.01	0.04	0.1	0.4
lag_Spa	2,284.2	1,369.8	0.0	1,273.6	2,151.2	3,289.0	5,590.9
lag_log_Pop	11.0	0.2	10.6	10.9	11.0	11.1	11.4

5 Results and discussion

In our analysis, we decided to proceed from the simplest model, including only the main variables of our interest, to the more complex ones, which should deal with all forms of endogeneity for the two essential factors. Table 3 shows the results of the baseline OLS models involving the variables of cultural and natural heritage.

Table 5: Basic OLS models

	Dependent variable:		
	Visitors	VpC	log_VpC
	(1)	(2)	(3)
Cult	173.047**	1.057	
	(78.755)	(1.008)	
log_Cult			0.562***
			(0.149)
Nature	9,831.237***	108.457***	0.049**
	(1,636.226)	(20.939)	(0.021)
Constant	−3,034.060	245.627	3.380***
	(14,034.650)	(179.606)	(0.659)
Observations	71	71	71
R^2	0.422	0.321	0.305
Adjusted R^2	0.405	0.301	0.284
Residual Std. Error (df = 68)	73,975.310	946.686	0.925
F Statistic (df = 2; 68)	24.825***	16.069***	14.887***
Note:	*p<0.1; **p<0.05; ***p<0.01		
Standard errors in parentheses			

The column (1) shows a very significant positive effect of the natural reserves and monuments, and also a positive effect of cultural monuments.

Our motivation was inspired by the fact that many locations dispose of high amounts of those attractions, but only some are popular tourist destinations. When we looked at the residuals, we discovered vast differences between districts with similar amounts of heritage sites.

We start with the variable representing cultural monuments. Since this type of heritage is linked to the history of people living in an area, and many of the current more populous towns and cities used to be regional centers also in the past (such as Košice and Nitra), it is logical that a large fraction of existing historical monuments is preserved in those locations. On the other hand, some former centers lost their importance over time, but despite their small population, they belong to the districts with the highest number of monuments (such as Banská Štiavnica and Levoča). Considering that big cities also offer other attractions, which cannot be found in smaller towns, controlling for the population as a proxy for all those variables, we might obtain more realistic estimates.

In column (2), we included the number of visitors per capita (multiplied by 1000, more about the variable can be found in section 4) as the dependent variable. It is visible that the estimate of the variable Cult is no more significant. However, this fall in significance could be expected because the more populous districts must, in absolute numbers, attract a higher number of visitors than the smaller ones with the same amount of cultural heritage in order to achieve the same level of VpC. We overcome this problem in the unrestricted models, mainly by including the variable representing the population.

Two districts with by far the largest positive residuals in the 1st model (in the 2nd, they still count in the top three) were Poprad and Liptovský Mikuláš - both located in the area of the Tatra mountains. At the same time, both belong to the three districts with the largest sum of natural reserves and monuments (26 and 22). With 24 reserves and monuments, the district Rožňava is in second place. However, Rožňava is also the largest negative residual in both models 1 and 2. It is clear that there must exist something else, which links Poprad and Liptovský Mikuláš, but not Rožňava and other districts. At the same time, Poprad and Liptovský Mikuláš are districts with also by far the largest number of visitors and visitors per capita. It can be assumed that this is the reason for the significant effect of Nature, even though there are huge differences between locations. Applying a log-form of the dependent variable can help to decrease the effect of large values on the estimates and make the residuals have a distribution nearer to normal.

As expected, using log_VpC instead of VpC as a dependent variable made the significance of the variable Nature smaller. However, the extreme values of the most visited districts in models 1 and 2 also probably affected the variable Cult. As shown in Figure 2, after using the logarithm of VpC, the positive relationship between the dependent variable and the amount of cultural and natural heritage becomes visible. Also, the shape of the curve became noticeable, which implied the use of logarithmic transformation also for the variable Cult, as presented in column (3). Compared to model 1, in model 3, the effects of explanatory variables are still positive, but they are more realistic. Therefore, the lower R^2 is not a surprise in this case. In Table 4, we show the models with additional independent variables.

Table 6: Adjusted OLS models - 1

	Dependent variable:	
	log_VpC	
	(4)	(5)
log_Cult	0.677***	0.717***
	(0.132)	(0.137)
Nature	0.023	0.039
	(0.022)	(0.024)
Ski	1.461***	0.828
	(0.507)	(0.546)
Spa	0.0001***	0.0001***
	(0.00002)	(0.00002)
log_Pop	−0.291*	−0.362**
	(0.162)	(0.171)
BA_Dist	−0.003***	
	(0.001)	
Unempl		−0.112***
		(0.028)
Constant	6.577***	7.133***
	(1.613)	(1.692)
Observations	71	71
R^2	0.615	0.604
Adjusted R^2	0.579	0.567
Residual Std. Error (df = 64)	0.709	0.719
F Statistic (df = 6; 64)	17.023***	16.272***
Note:	*p<0.1; **p<0.05; ***p<0.01	
Standard errors in parentheses		

Now, we can look at other sources of endogeneity. We mentioned districts Poprad and Liptovský Mikuláš and their role in the positive effect of the natural heritage. Both regions belong to the most popular destinations in Slovakia, and tourism infrastructure is undoubtedly the most developed in those regions. However, the tourism infrastructure is usually only a secondary product that has been developed because of the attractive places and attractions there. The common factor for both districts is the mountains, especially the Low and High Tatras. As the highest mountains in Slovakia, they offer tourists not only hiking opportunities during the main holiday season in summer but also the largest winter resorts are located there. In our opinion, the possibility to attract masses of tourists throughout the whole year, together with a wide range of offered activities (such as wellness and aquaparks, adrenaline sports, nightlife, et cetera) and other parts of tourism infrastructure are all the reasons of their success.

Nevertheless, the crucial element of it is the high-quality ski resorts. Therefore, we decided to use the variable representing the quality of ski resorts as a proxy for all the tourism infrastructure related to winter resorts. Models in columns (4) and (5) show that the variable Nature is no longer significant.

When we look at the residuals of model 3 once again, we can see that the three districts with the highest positive residuals are Turčianske Teplice, Krupina, and Piešťany. A common characteristic for them is the presence of important spa resorts (Turčianske Teplice, Dudince, and Piešťany). Out of the top 13 positive residuals, 7 have popular spas (in the lower half of the table are only three). Even though we did not use the variable representing overnight stays as a dependent, a variable describing spa resorts still should be included in the model. A potential reason could be again the whole year tourism. Such variable in our new models is Spa, representing the importance of popular spa resorts.

Despite we included the variable representing population in the dependent variable, using the logarithmic form of this variable as an independent one still provides a piece of important information, as we wrote for model 2. In the results of models 4 and 5, we can see significant negative effects of this variable. This is a logical result, as was mentioned because having VpC as the dependent variable, the larger districts are disadvantaged compared to smaller in terms of the absolute numbers of visitors. At the same time, the benefits for tourists resulting from larger opportunities in big cities are eliminated by having the dependent variable measured per capita. Therefore, the estimates for our two main explanatory variables should not be biased because of the number of inhabitants.

The last variable we included in the model, the distance of a district from Bratislava, might be relevant for two reasons. Firstly, as mentioned in the Section 4, the capital and largest city might impact the number of visitors in neighboring districts. In this case, it would be enough to study the effect of distance only on the regions located close to Bratislava. Nevertheless, we found in the data that the number of visitors generally starts to decrease after some distance. If we look at the residuals again, we find that out of the 17 biggest negative residuals, 15 are located in eastern and south-eastern Slovakia. Many of those districts belong to the least-developed districts in Slovakia. The worse economic situation in those regions is without a doubt caused and causes various problems, which are surely reflected also in insufficient tourism infrastructure and tourism performance.

As a proxy variable for underdevelopment and linked problems can be used unemployment. There is a high correlation between variables Unempl and BA_Dist: 0.74. In models 4 and 5, we have tried to compare both variables. Model 4 slightly outperforms model 5, at least in terms of its fit. An interesting difference is that in model 5, neither Nature nor Ski is significant. There is a good reason for that: The district with the second highest number of natural heritage, Rožňava, was the district with the third largest unemployment in 2018, 13.92%. However, there is a problem with causality in this case: Is Rožňava's undeveloped tourism potential caused by its economic and social situation, or could this district have been more developed if it was more invested in its tourism products? Clearly, there is no single easy answer, and the problem is in both directions. Even though there are some differences between variables Unempl and BA_dist, and including both might make the model a bit more precise, we want to include only the most essential variables because of our small observation set. We continue only with the variable BA_dist in further models because unemployment has a relatively high correlation with the distance from Bratislava, its causality is not entirely clear, and it has worse explanation power.

Model 4 is our final OLS model, so we test its assumptions. Based on the RESET test, we cannot reject on the 95% confidence interval that our model suffers from none functional misspecifications up to the third order (p-value = 0.485). Similarly, the VIF-test showed no values higher than 5 (not even 3), so our model does not have a multicollinearity problem. According to the Breusch-Pagan and White tests, we cannot reject homoscedasticity in the model on the 95% confidence level (p-value is 0.27 and 0.16, respectively). Regarding normality, we found a distribution of residuals relatively close to normal. To be sure, we also applied the Shapiro test, and according to it, we cannot reject the normality of residuals on the 95% level of confidence (p-value = 0.16).

We have not discovered any further suspicious common characteristics in residuals of model 4. Before continuing with the spatial models, we look at one spatial interaction, which does not fully fit into this category. We mentioned before that the capital Bratislava might affect the results even though it is not directly involved as an observation. There are two possible ways.

Firstly, it can negatively affect its neighboring regions because it might steal their potential overnight visitors. The benefits of the big city, such as nightlife or shopping, have already been mentioned many times. Another reason might be that Bratislava is a much better point of departure if someone wants to visit the surrounding region. There are usually direct routes and trains from the capital city to all nearby towns, but they might not exist between towns on each other.

The positive effect has also been partially mentioned. Being close to the capital means better job opportunities for locals than, for example, in eastern Slovakia. This might naturally increase the quality of life in those regions and attract other people to live there, leading to new investments into local infrastructure and possibly bringing new attractions for locals and tourists. Bratislava might not only steal tourists from the vicinity, but the effect can also be the opposite. Some tourists might prefer sleeping outside the noisy city and might decide, for example, for romantic accommodation in a nearby wine village.

We examined the described effect using two methods. The first one is more intuitive and means including a variable of an inverse distance from Bratislava to the districts nearby truncated by 50km (not bounded variable of distance is already present in the model) - presented in Table 5 as a column (6). The other possibility is to include the quadratic form of the variable BA_dist because we know that the districts in the east have generally smaller numbers of VpC. If we also find the smaller values than predicted for the districts near Bratislava, we would have a curve described by the quadratic polynomial - column (7). In column (8), we repeated model 7 with the variable BA_dist demeaned and dived by 100.

In model 6, we do not see any significant effects of the variable rep-

Table 7: Adjusted OLS models - 2

	Dependent variable:		
	log_VpC		
	(6)	(7)	(8)
log_Cult	0.671***	0.634***	0.634***
	(0.132)	(0.133)	(0.133)
Nature	0.021	0.019	0.019
	(0.022)	(0.022)	(0.022)
Ski	1.511***	1.286**	1.286**
	(0.509)	(0.513)	(0.513)
Spa	0.0001***	0.0001***	0.0001***
	(0.00002)	(0.00002)	(0.00002)
log_Pop	−0.301*	−0.225	−0.225
	(0.162)	(0.166)	(0.166)
BA_Dist	−0.003***	0.002	
	(0.001)	(0.003)	
BA_Dist_u50	12.495		
	(11.736)		
BA_Dist_2		−0.00001	
		(0.00001)	
BA_Dist_cent			−0.237***
			(0.079)
BA_Dist_cent_2			−0.095
			(0.060)
Constant	6.615***	5.615***	5.564***
	(1.611)	(1.706)	(1.568)
Observations	71	71	71
R^2	0.622	0.630	0.630
Adjusted R^2	0.580	0.588	0.588
Residual Std. Error (df = 63)	0.709	0.701	0.701
F Statistic (df = 7; 63)	14.784***	15.294***	15.294***

Note: *p<0.1; **p<0.05; ***p<0.01

Standard errors in parentheses

resenting closeness to Bratislava. In model 7, the estimate of BA_Dist is positive, and the estimate of its squared form is negative, which is a sign of a quadratic curve we are looking for, but both estimates are not significant. The centered and transformed form in model 8 provides better insight because we no longer have a problem with high collinearity between the two variables, and the standard errors are more readable than in model 7. We see a significant negative effect of the linear variable, but the quadratic form is not significant on the 90% confidence level. However, it is already significant on the 88% confidence interval, which is not enough to make any valid conclusions, but it is good to keep in mind that such an effect might exist.

5.1 Spatial effects

We open the second part of the analysis with testing for spatial interactions in model 4. We start by testing for the presence of spatial error and lag dependences, and we continue by building the SLX model to prove or reject the autocorrelation in independent variables.

For the first two effects, we use the Lagrange multiplier tests. Both tests were repeated for the row-normalized versions of the adjacency (Table 6) and inverse distance weighting matrices (Table 7) because we think these two types might be the most relevant for our case, as described in the methodology.

Table 8: LM tests for spatial dependence - adjacency matrix

	statistic	parameter	p.value
Spatial error dependence	0.582158371347316	df = 1	0.445
Spatial lag dependence	0.644948306408534	df = 1	0.422

Table 9: LM tests for spatial dependence - inverse-distance matrix

	statistic	parameter	p.value
Spatial error dependence	0.69486589974692	df = 1	0.405
Spatial lag dependence	3.29843249641573	df = 1	0.069

None of the tests indicates that we should reject the null hypothesis of no spatial dependence at the 95% confidence interval. However, the test for spatial lag dependence in Table 7 is not far away from it. We should remember this for the interpretation.

Table 10 shows the SLX model which we made to find out whether the independent variables' spatial lags might impact the two variables of our primary interest. We did not include the lag variable of BA_Dist, because neighboring and near regions have almost the same distance from the capital as the specific districts. Another thing to remember is the number of observations, which is only 71, and we have 11 independent variables, so we might face a high risk of overfitting the model. Also, the VIF test shows high values between some lags. For both reasons, we present the models only to look at the lags of cultural and natural heritage sites.

Table 10: Full SLX models

	Dependent variable:	
	log_VpC	
	(9 - Adjacency matrix)	(10 - Inverse-dist. matrix)
log_Cult	0.552***	0.507***
	(0.163)	(0.147)
Nature	−0.001	0.009
	(0.022)	(0.021)
Ski	1.088**	1.079**
	(0.502)	(0.502)
Spa	0.0001***	0.0001***
	(0.00002)	(0.00002)
log_Pop	−0.087	−0.080
	(0.184)	(0.171)
BA_Dist	−0.004***	−0.004***
	(0.001)	(0.001)
lag_log_Cult	0.111	0.153
	(0.242)	(0.346)
lag_Nature	0.023	−0.096
	(0.039)	(0.100)
lag_Ski	1.667*	4.419**
	(0.848)	(2.073)
lag_Spa	−0.0001**	−0.0001
	(0.00004)	(0.0001)
lag_log_Pop	−0.733**	−1.431***
	(0.354)	(0.486)
Constant	12.726***	20.618***
	(3.473)	(5.006)
Observations	71	71
R^2	0.692	0.696
Adjusted R^2	0.635	0.639
Residual Std. Error (df = 59)	0.660	0.656
F Statistic (df = 11; 59)	12.072***	12.270***
Note:		*p<0.1; **p<0.05; ***p<0.01
Standard errors in parentheses		

The estimates and variances in models in columns (9) and (10) are relatively similar with a few exceptions. The important observations from these models are that the estimates of the main explanatory variables changed only slightly, and their lags seem to be non-significant.

We start with the variable lag_log_Cult. The VIF test shows values under 3 for both models, so it is unlikely that it is insignificant because of high multicollinearity. To discover why the effect of the variable seems insignificant, we look at its values more closely. The numbers do not have a form of normal distribution but rather a distribution with two peaks (more visible in the inverse-distance matrix case). The part/curve on the left involves the majority of data. The smaller one on the right is characterized by observations from approximately the same location: the historical region of Spiš and districts nearby. This area generally has many cultural monuments, so the districts there positively affect each other's lag_log_Cult. However, some of those districts, like Poprad, have one of the highest values of log_VpC. The others, like Gelnica, belong to districts with the lowest. Therefore, this variable is very heterogenous and cannot have a significant effect.

The reason for the insignificance of lag_Nature is probably similar to the one for variable Nature, but also variable lag_log_Cult. Among the variables with the largest values of lag_Nature are on one side districts near the Tatra mountains, such as Ružomberok or Kežmarok. On the other side, there are also districts neighboring Rožňava and caves of the Slovak

Karst National park, such as Revúca. At the same time, both regions are geographically near each other (Poprad and Rožňava actually border), which causes districts in this area to have similar values of lag variables. Therefore, we probably obtained insignificant estimates also for this variable.

5.2 Robustness check

As a robustness check, we planned to run regression on the data for 2017 and 2019 to show that our results are not affected by some random events or shocks occurring in 2018 (as would be the case in 2020 and 2021 because of the pandemics), or by missing values. However, 2017 also contained some missing values in the dependent variable, so we performed the test only for 2019. The only missing values for this year were districts Košice II and III, which could be easily overcome, similarly to 2018.

As was written in Section 4, the numbers of the natural and cultural heritage change usually only slightly in such a short time. The same can be said about variables Pop, Ski, and Spa (for the proxy variable Ski, we were even forced to use values from a different year due to data availability). The distance from Bratislava does not change in time at all. Therefore, we used different values only for the dependent variable - VpC (Statistical Office of the Slovak Republic 2019). The results are presented in column (12) of Table 11, and column (11) shows the original final model for 2018. The dependent variable log_VpC_19 was created exactly the same way as log_VpC.

Table 11: Final OLS model and model for 2019

	Dependent variable:	
	log_VpC	log_VpC_19
	(11)	(12)
log_Cult	0.677***	0.654***
	(0.132)	(0.130)
Nature	0.023	0.019
	(0.022)	(0.022)
Ski	1.461***	1.563***
	(0.507)	(0.498)
Spa	0.0001***	0.0001***
	(0.00002)	(0.00002)
log_Pop	−0.291*	−0.316*
	(0.162)	(0.159)
BA_Dist	−0.003***	−0.003***
	(0.001)	(0.001)
Constant	6.577***	7.102***
	(1.613)	(1.585)
Observations	71	71
R^2	0.615	0.611
Adjusted R^2	0.579	0.575
Residual Std. Error (df = 64)	0.709	0.697
F Statistic (df = 6; 64)	17.023***	16.771***
Note:	*p<0.1; **p<0.05; ***p<0.01	
Standard errors in parentheses		

We can see that the estimates and their significance changed only a little. The same holds for the R^2. Therefore, we can conclude that the model and its results are robust enough to work at least also for 2019, but since our variables do not change in time very much, it would likely hold for other pre-pandemic years.

6 Conclusion and limitations

In this thesis, we analyzed the impacts of the cultural and natural heritage sites on the attractiveness of Slovak districts. A common opinion in academic research is that those monuments and places belong to the most crucial attractions, according to which visitors choose a certain destination. Despite this logical assumption, we found vast differences in the number of visitors to the districts with a similar abundance of cultural and natural monuments. We tried to answer whether it is really the natural and historical beauties, which attract visitors to spend time in Slovak hotels, or whether it is just a desire of local tourism planners that this is the main factor tourists appreciate in Slovakia.

We analyzed the 71 observations representing all Slovak districts except all five separate districts of Bratislava, and we merged all four districts of Košice into one observation. Bratislava affected our variables as an outlier, and we argued that it is more reasonable to include Košice as one city. We used the annual data for 2018 because it was the last pre-pandemic year, for which we could compare the data with the preceding and consecutive year without dealing with pandemic shocks. Since all the variables changed relatively slightly throughout the years, we could complete the missing values without concerns about their correctness.

As the dependent variable, we used the number of arrivals to accommodation establishments of tourism, which is a typical variable for tourism demand representing the perceived attractiveness. The independent variables of our interest were the number of national cultural monuments and the number of national nature monuments and reserves, which are common in the tourism supply studies focusing on attractiveness phenomenon from the perspective of existing tourism resources.

In the baseline OLS models, including only the main independent variables, we showed that both factors seem to have a positive impact on tourism visits because estimates of cultural and natural heritage were positive and significant. However, in the consecutive analysis, we found multiple sources of endogeneity for these models.

In the following OLS models, we applied some transformations to the data and included additional explanatory variables. The relatively small number of observations allowed us to study the residuals of the baseline model in detail. We found that the vast differences in tourism arrivals between the observations with a similar amount of heritage sites can be explained by finding common characteristics of the largest positive and negative residuals. Four additional crucial independent variables from the tourism supply field were identified using this approach: factors representing the quality and importance of ski and spa resorts, the distance from the capital Bratislava, and population. We also divided the dependent variable per capita, and we applied the logarithmic transformation for some variables.

In the results of the extended OLS models, the estimate of cultural heritage remained positive and significant, but the natural sites lost their significance. We found out that the positive estimates of the second mentioned variable in the baseline model were affected mainly by the districts in the area of the Tatra mountains, which belong in terms of the number of visitors and visitors per capita to the most attractive locations in Slovakia. The districts with many historical buildings are generally not visited to such a great extent, at least in terms of the whole year numbers. However, we found clear evidence of the positive impacts of the number of cultural monuments.

We tested the robustness of our results on the dependent variable for 2019 (the preceding year, 2017, involved some missing values). As expected, we found only tiny differences compared to the original model. The tourism demand is limited, among others, by the capacities of tourism accommodation establishments, and in case no shocks occur, such as current pandemics or organizing some massive events, in our opinion, there is no reason to assume huge changes in tourism demand in the short-run.

Since our analysis operates with geographical units, the spatial interactions between the districts represent an essential part of the analysis. The first effect, which could also be observed in residuals and for which we did not need to apply any specific spatial econometric models, was the distance from Bratislava. We examined it for multiple reasons. Firstly, we did not include this observation in our dataset, but we found arguments that as the capital, it might still influence the visits of neighboring regions. The second reason is the huge Slovakia's regional disparities, which are visible mainly in the west-east country division. These differences can be found across many economic and non-economic areas. We also tried using unemployment as a proxy for all those differences, but we found its high correlation with the distance from Bratislava, which explained the data better. We found a significant negative impact of a higher distance from Bratislava. This effect could also explain why we found the positive effect of cultural heritage, even though many of the districts with the largest amounts of historical monuments are located in eastern Slovakia but belong not to the most visited locations from the perspective of the whole country.

We also discovered some indications that Bratislava might have a negative effect on districts in its vicinity, which could be interpreted as stealing the potential hotel guests from the nearby districts. The whole effect of Bratislava might thus have a quadratic form. However, the impact of the squared variable was significant only on the 88% confidence level.

In the following, we tested for spatial lag and spatial lag errors using the Lagrange multiplier tests for both adjacency and inverse distance matrix. We could not reject the null hypothesis of no spatial autocorrelation for all tests on the 95% confidence level. However, the test for the spatial lag dependence performed on the inverse distance weighting matrix was not far from the threshold, which might indicate that an analysis of some related topic or variables might show the need to include the spatial lags of the dependent variable in the model.

As the last step, we estimated the SLX model to determine whether we find some effects of spatial lags of the cultural and natural heritage. We were aware of a potential problem of this model resulting from the fact that the estimation of spatial lags of the independent variables almost doubled their number. It could cause a problem with model overfitting. Therefore, we focused only on interpreting the values of spatial lags of both main variables. We found out that Slovakia's geography did not allow any significant effects of spatial lags of cultural and mainly natural heritage. The reason is that the most visited destinations near the High and Low Tatra mountains are bordering regions with similarly rich places on those valuable sites, but due to various problems, they belong to the least visited. Despite geographical proximity, both large regions are located in the mountains, and traveling from one to another is relatively difficult. In our opinion, this is the reason why they behave as if they were not neighbors.

Our models also have some limitations. Firstly, it is the relatively small number of observations. It could be theoretically solved by using smaller spatial units. However, there is a problem with little data available for that size. Secondly, it is the geography of Slovakia. The country is not an island, and it is possible that the regions on the other side of the border might also affect the results. Also, some large cities are within a few hours distance (Budapest, Krakow, Prague, Vienna). However, also studying the spatial effects of neighboring countries would make the scope of the analysis much larger. It is also connected to another problem: we did not differentiate between domestic and foreign visitors. The tourists' home location might involve the model in various ways, such as proximity to a destination, marketing, price differences, etc. However, also studying the variables typical for tourism demand would make the number of independent variables even larger.

Another thing is that we presented only one way to measure cultural and natural heritage, which we think represents both qualitative and quantitative aspects of the phenomena and which data were available. Academic literature also used different ways, which might provide helpful information. Despite this, they often contain only one aspect of information: quality or quantity, and it is impossible to find data for most of them. An extension of the analysis might also include the time dimension in the models. Especially interesting might be the effect of the pandemics. Lastly, we found a paper arguing that the tourism system is by its nature extremely complex and nonlinear. We also found some examples of it in our models when the logarithmic transformations and quadratic forms seem more appropriate. Analyzing the high-complex nonlinear models is above the scope of the bachelor thesis.

As far as we know, no academic papers are calling into question that the natural and cultural heritage is the main sources of tourism demand in Slovakia. When we chose this topic, we expected more to find a non-significant effect of cultural heritage than the natural one, probably because tourism marketing often emphasizes the natural beauties of Slovakia. Nevertheless, as we showed, the visitors' attention is often oriented only toward the Tatra region.

Potential implications from our results might be used by tourism planners, marketers, and other people employed in the travel and tourism industry. The first recommendation is to focus more on promoting other than high-mountains regions when talking about locations rich in natural monuments. Many of those beautiful natural sceneries are located in economically underdeveloped regions and are not affected by mass tourism. We believe the development of sustainable tourism destinations might have huge potential there.

At the same time, the tourism potential of cultural sites in Slovakia is sometimes unjustly forgotten compared to the natural sites. However, there clearly is a demand for this type of destination. The attractiveness of those locations can be increased not only by intensive marketing, but in our opinion, if the demand should be satisfied, we need investments in the infrastructure. The reason is that the potential of many of those locations is diminished by a relatively large distance from Bratislava, which is generally connected to higher unemployment and a worse state of local development.

References

Assaf, A. G. & Tsionas, M. G. (2019), 'A review of research into performance modeling in tourism research - launching the annals of tourism research curated collection on performance modeling in tourism research', *Annals of Tourism Research* **76**, 266–277.

Backman, S. J., Uysal, M., Backman, K. et al. (1991), 'Regional analysis of tourism resources.', *Annals of Tourism research* **18**(2), 323–327.

Baggio, R. (2008), 'Symptoms of complexity in a tourism system', *Tourism Analysis* **13**(1), 1–20.

Balli, F., Balli, H. O. & Louis, R. J. (2016), 'The impacts of immigrants and institutions on bilateral tourism flows', *Tourism Management* **52**, 221–229.

Barros, C. P., Botti, L., Peypoch, N., Robinot, E., Solonandrasana, B. & A., G. A. (2011), 'Performance of french destinations: Tourism attraction perspectives', *Tourism Management* **32**(1), 141–146.

Barros, C. P. & Dieke, P. U. (2008), 'Technical efficiency of african hotels', *International Journal of Hospitality Management* **27**(3), 438–447.

Bassil, C. (2014), 'The effect of terrorism on tourism demand in the middle east', *Peace Economics, Peace Science and Public Policy* **20**(4), 669–684.

Bauhoff, S. (2005), 'Social science statistics blog: Spatial error'. Harvard

University, Accessed: 18-7-2022.
URL: *https://blogs.iq.harvard.edu/spatial_error_1*

Benito, B., Solana, J. & López, P. (2014), 'Determinants of spanish regions' tourism performance: A two-stage, double-bootstrap data envelopment analysis', *Tourism Economics* **20**(5), 987–1012.

Bivand, R. (2022), 'R packages for analyzing spatial data: A comparative case study with areal data', *Geographical Analysis* **54**(3), 488–518.

Bučeková, I., Erbert, L. & Klobučník, M. (2019), 'Klasifikácia lyžiarskych stredísk na slovensku', *Geografický časopis* **71**, 363–382.

Buhalis, D. (2001), 'Tourism in greece: Strategic analysis and challenges', *Current Issues in Tourism* **4**(5), 440–480.

Calderwood, L. U. & Soshkin, M. (2019), 'The travel and tourism competitiveness report 2019', World Economic Forum.

Capital City of Slovakia - Bratislava (n.d.), 'Bratislava 72 hours city'. Accessed: 23-07-2022.
URL: *https://www.visitbratislava.com/*

Cerisola, S. (2019), 'A new perspective on the cultural heritage–development nexus: The role of creativity', *Journal of Cultural Economics* **43**(1), 21–56.

Cha, S. & Uysal, M. (1995), 'Regional analysis of tourism resources: A case study of korea', *Journal of Hospitality & Leisure Marketing* **2**(3), 61–74.

Chaabouni, S. (2019), 'China's regional tourism efficiency: A two-stage double bootstrap data envelopment analysis', *Journal of Destination Marketing Management* **11**, 183–191.

Chi, G. & Zhu, J. (2019), *Spatial Regression Models for the Social Sciences*, Vol. 14, SAGE Publications.

Cho, V. (2010), 'A study of the non-economic determinants in tourism demand', *International Journal of Tourism Research* **12**(4), 307–320.

Council of the European Union (2014), 'Conclusions on cultural heritage as a strategic resource for a sustainable europe: Education, youth, culture and sport'. Accessed: 10-5-2020.

URL: *https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/educ/142705.pdf*

Crouch, G. I. & Ritchie, J. (1999), 'Tourism, competitiveness, and societal prosperity', *Journal of Business Research* **44**(3), 137–152.

Cuccia, T., Guccio, C. & Rizzo, I. (2016), 'The effects of unesco world heritage list inscription on tourism destinations performance in italian regions', *Economic Modelling* **53**, 494–508.

de Castro, E. V., Souza, T. B. & Thapa, B. (2015), 'Determinants of tourism attractiveness in the national parks of brazil', *Parks* **21**(2), 51–62.

Dritsakis, N. (2004), 'Cointegration analysis of german and british tourism demand for greece', *Tourism management* **25**(1), 111–119.

Elhorst, J. P. (2014), *Spatial econometrics from cross-sectional data to spatial panels*, SpringerBriefs in Regional Science.

Elhorst, P., Vega, S. H. et al. (2013), On spatial econometric models, spillover effects, and w, *in* 'ERSA conference papers', number ersa13p222, European Regional Science Association.

European Commission and Directorate-General for Research and Innovation (2015), *Getting cultural heritage to work for Europe : report of the Horizon 2020 expert group on cultural heritage*, Publications Office.

Farid, S. M. (2015), 'Tourism management in world heritage sites and its impact on economic development in mali and ethiopia', *Procedia - Social and Behavioral Sciences* **211**, 595–604. 2nd Global Conference on Business and Social Sciences (GCBSS-2015) on "Multidisciplinary Perspectives on Management and Society", 17- 18 September, 2015, Bali, Indonesia.

Formica, S. (2000), Destination attractiveness as a function of supply and demand interaction, PhD thesis, Virginia Polytechnic Institute and State University.

Formica, S. & Uysal, M. (2006), 'Destination attractiveness based on supply and demand evaluations: An analytical framework', *Journal of Travel Research* **44**(4), 418–430.

Gúčik, M., Kvasnová, D. & Pančíková, K. (2016), 'Medical spa versus health tourism', *Acta academica karviniensia* **16**(2), 5–15.

Gunn, C. A. (1972), *Vacationscape: Designing Tourist Regions*, University of Texas, Austin: Bureau of Business Research.

Gyr, U. (2010), *Europäische Geschichte Online : EGO : The history of tourism : structures on the path to modernity*, Inst. f. Europ. Geschichte, Mainz. Accessed: 06-07-2022.

URL: *http://www.ieg-ego.eu/gyru-2010-en*

Halleck Vega, S. & Elhorst, J. P. (2015), 'The slx model', *Journal of Regional Science* **55**(3), 339–363.

Hu, Y. & Ritchie, J. B. (1993), 'Measuring destination attractiveness: A contextual approach', *Journal of travel research* **32**(2), 25–34.

Iatu, C. & Bulai, M. (2011), 'New approach in evaluating tourism attractiveness in the region of moldavia (romania)', *International Journal of Energy and Environment* **5**(2), 165–174.

IUCN (2014), The benefits of natural world heritage : identifying and assessing ecosystem services and benefits provided by the world's most iconic natural places, Technical report, IUCN, Federal Agency for Nature Conservation (BfN), DE, IUCN, World Heritage Programme, UNEP, WCMC. ISBN: 978-2-8317-1694-7.

Jafari, J. (1974), 'The components and nature of tourism: The tourism market basket of goods and services', *Annals of tourism research* **1**(3), 73–89.

Kaur, J. (1981), 'Methodological approach to scenic resource assessment', *Tourism Recreation Research* **6**(1), 19–22.

Kelejian, H. & Piras, G. (2017), *Spatial econometrics*, Academic Press - an imprint of Elsevier.

Koens, K., Postma, A. & Papp, B. (2018), 'Is overtourism overused? understanding the impact of tourism in a city context', *Sustainability* **10**(12), 4384.

Kotrč, M. (2020), The impact of structural funds and cohesion fund on convergence: The case of slovak districts, diplomová práce, Univerzita Karlova, Fakulta sociálních věd, Institut ekonomických studií, Praha. Vedoucí práce Šťastná, Lenka.

Kozak, N., Uysal, M. & Birkan, I. (2008), 'An analysis of cities based on tourism supply and climatic conditions in turkey', *Tourism Geographies* **10**(1), 81–97.

Kusni, A., Kadir, N. & Nayan, S. (2013), 'International tourism demand in malaysia by tourists from oecd countries: A panel data econometric analysis', *Procedia Economics and Finance* **7**, 28–34.

Lew, A. A. (1987), 'A framework of tourist attraction research', *Annals of tourism research* **14**(4), 553–575.

Li, G., Song, H. & Witt, S. F. (2005), 'Recent developments in econometric modeling and forecasting', *Journal of Travel Research* **44**(1), 82–99.

Lim, C. (1997), 'Review of international tourism demand models', *Annals of tourism research* **24**(4), 835–849.

Lim, C. & McAleer, M. (2002), 'A cointegration analysis of annual tourism demand by malaysia for australia', *Mathematics and Computers in Simulation* **59**(1-3), 197–205.

Lovingood Jr, P. E. & Mitchell, L. E. (1989), 'A regional analysis of south carolina tourism', *Annals of Tourism research* **16**(3), 301–317.

Matijová, M., Onuferová, E., Rigelský, M. & Stanko, V. (2019), 'Impact of selected indicators of tourism capacity and performance in the context of the unemployment rate in slovakia', *Journal of tourism and services* **10**(19), 1–23.

Mayo, E. J., Jarvis, L. P. et al. (1981), *The psychology of leisure travel. Effective marketing and selling of travel services.*, CBI Publishing Company, Inc.

McKercher, B. (2016), 'Towards a taxonomy of tourism products', *Tourism Management* **54**, 196–208.

Moore, W. R. (2010), 'The impact of climate change on caribbean tourism demand', *Current Issues in Tourism* **13**(5), 495–505.

Morley, C. L. (1992), 'A microeconomic theory of international tourism demand', *Annals of tourism research* **19**(2), 250–267.

Nyberg, L. (1995), Determinants of the attractiveness of a tourism region, *in* S. Witt & L. Moutinho, eds, 'Tourism Marketing and Management Handbook', Hertfordshire: Prentice Hall, pp. 29–38.

O'Hagan, J. W. & Harrison, M. (1984), 'Market shares of us tourist expenditure in europe: an econometric analysis', *Applied economics* **16**(6), 919–931.

Pompurova, K. & Simockova, I. (2014), 'Destination attractiveness of slovakia: perspectives of demand from major tourism source markets', *E+ M Ekonomie a Management* **17**(3), 62–74.

Romao, J., Guerreiro, J. & Rodrigues, P. M. (2013), Territorial differentiation, competitiveness and sustainability of tourism, *in* 'Quantitative Methods in Tourism Economics', Springer, pp. 271–285.

Romão, J. (2015), Culture Or Nature: A Space-Time Analysis On The Determinants Of Tourism Demand In European Regions, Spatial and Organizational Dynamics Discussion Papers 2015-3, CIEO-Research Centre for Spatial and Organizational Dynamics, University of Algarve.

Shang, J.-K., Wang, F.-C. & Hung, W.-T. (2010), 'A stochastic dea study of hotel efficiency', *Applied Economics* **42**(19), 2505–2518.

Shoval, N. & Raveh, A. (2004), 'Categorization of tourist attractions and the modeling of tourist cities: based on the co-plot method of multivariate analysis', *Tourism Management* **25**(6), 741–750.

Siikamäki, P., Kangas, K., Paasivaara, A. & Schroderus, S. (2015), 'Biodiversity attracts visitors to national parks', *Biodiversity and conservation* **24**(10), 2521–2534.

Sinclair, M. T., Blake, A. & Sugiyarto, G. (2003), The economics of tourism, *in* C. P. Cooper, ed., 'Classic reviews in tourism', Clevedon: Channel View Publications, pp. 22–54.

SITA (2020), 'V tatrách padol rekord v návštevnosti, najviac lákalo popradské pleso', Korzár SME.
URL: *https://spis.korzar.sme.sk/c/22464507/inventura-turistov-v-tatrach-padol-40-rocny-rekord-v-navstevnosti.html*

Slovakia Travel (n.d.), 'Národný portál cestovného ruchu slovenska'.
URL: *https://slovakia.travel/en*

Smith, S. L. (1987), 'Regional analysis of tourism resources', *Annals of tourism research* **14**(2), 254–273.

Smith, S. L. (1988), 'Defining tourism a supply-side view', *Annals of tourism research* **15**(2), 179–190.

Song, H. & Li, G. (2008), 'Tourism demand modelling and forecasting—a review of recent research', *Tourism management* **29**(2), 203–220.

Song, H., Li, G., Witt, S. F. & Fei, B. (2010), 'Tourism demand modelling and forecasting: how should demand be measured?', *Tourism economics* **16**(1), 63–81.

Spotts, D. M. (1997), 'Regional analysis of tourism resources for marketing purposes', *Journal of Travel Research* **35**(3), 3–15.

Statistical Office of the Slovak Republic (2018*a*), 'Miera evidovanej nezamestnanosti [vbd_sk_win:pr3108rr]'. [Data file] Accessed: 25-09-2020.
URL: *http://datacube.statistics.sk/*

Statistical Office of the Slovak Republic (2018*b*), 'Počet obyvateľov podľa pohlavia [om7102rr]'. [Data file] Accessed: 18-06-2020.
URL: *http://datacube.statistics.sk/*

Statistical Office of the Slovak Republic (2018*c*), 'Štatistika ubytovacích zariadení [vbd_sk_win:cr3002rr]'. [Data file] Accessed: 25-09-2020.
URL: *http://datacube.statistics.sk/*

Statistical Office of the Slovak Republic (2019), 'Capacity and performances of accommodation establishments of tourism [vbd_sk_win:cr3002rr]'. [Data file] Accessed: 19-07-2020.
URL: *http://datacube.statistics.sk/*

Statistical Office of the Slovak Republic (2021), 'Slovenské cestovné kancelárie majú za sebou najzložitejší rok svojej existencie'. Accessed: 19-06-2022.
URL: *https://slovak.statistics.sk/wps/portal/ext/ab*

outus/office.activites/officeNews/vsetkyaktuality/60276d6f-a7ba-4585-81ab-3e70947789b4/!ut/p/z1/tVJNc4IwEP0tHjhCFvlI7C06HbGVg3askEsnIEKqEIQI9d83dnpoO7W1h-aQz_d2X_YtYihCrOKdyLkSsuJ7fY6Z_7TAMzIe2xQgvJ_AzJkFizFd2q49RI-IIZZWqlYFimXS8sJsd6aotibfKQP0JDqhTno98r3eGNC1mdp9vPBhiP2Nrxk44abrEc8kNk9MJ8MwcjEmo8Q9Z6lTsUHxVej1b7KZfoYLg4LmszfIZEoDF88ByHzqwYwGq-Vo4ThAnXfADzFirQFf1BDYaN2JrEerSjalrvTDH78YALpDTCSl1aelBZZNiE9c8G3Pc7E9Av_s3bAJJ2Gu43JVnE2RKPpSfhRdlU3HEs-HA6PabFmp7EWh6B_d1ulolThES2-ybdZkjXVsdDcWStXtjQEG9H1v5VLm-8xKZWnAd5RCtlrmZySqy5I4J303DW8dl8VEecmpp4PBK8ysc7c!/dz/d5/L2dBISEvZ0FBIS9nQSEh/

Statistical Office of the Slovak Republic (n.d.*a*), 'Capacity and performances of accommodation establishments of tourism - yearly data [cr2001rs]'. [Data file] Accessed: 19-06-2022.
URL: *http://datacube.statistics.sk/#!/view/en/VBD_SLOVSTAT/cr2001rs/v_cr2001rs_00_00_00_en*

Statistical Office of the Slovak Republic (n.d.*b*), 'Occupancy of accommodation establishments [cr3801mr]'. [Data file] Accessed: 19-06-2022.
URL: *http://datacube.statistics.sk/#!/view/en/vbd_sk_win2/cr3801mr/v_cr3801mr_00_00_00_en*

The Monuments Board of the Slovak Republic, M. S. (2020), 'Evidencia národných kultúrnych pamiatok na slovensku'. [Data file] Accessed: 13-02-2021.

URL: *http://www.pamiatky.sk/sk/page/evidencia-narodnych-kulturnych-pamiatok-na-slovensku*

Tisdell, C. A. (2013), Overview of tourism economics, *in* 'Handbook of tourism economics: analysis, new applications and case studies', World scientific, pp. 1–30.

TREND Analyses (2020), 'Trend top - rebríček najväčších kúpeľných spoločností (za rok 2019)'. [Data file] Accessed: 26-01-2022.

URL: *https://www.trend.sk/trend-archiv/trend-top-rebricek-najvacsich-kupelnych-spolocnosti*

Uysal, M. (1998), The determinants of tourism demand, *in* D. Ioannides & K. G. Debbage, eds, 'The economic geography of the tourist industry: A supply-side analysis', Psychology Press, pp. 79–95.

Vengesayi, S. (2003), A conceptual model of tourism destination competitiveness and attractiveness, *in* R. Kennedy, ed., 'Proceedings of the 2003 ANZMAC Conference', ANZMAC, pp. 637–647.

Vengesayi, S., Mavondo, F. T. & Reisinger, Y. (2009), 'Tourism destination attractiveness: Attractions, facilities, and people as predictors', *Tourism Analysis* **14**(5), 621–636.

Wooldridge, J. M. (2013), *Introductory Econometrics: A Modern Approach (Fifth Edition)*, South-Western Cengage Learning.

World Tourism Organization (2021), *The Economic Contribution of Tourism and the Impact of COVID-19, preliminary version*, UNWTO, Madrid.

WTTC - World Travel & Tourism Council (2020), 'Travel & tourism: Economic impact 2020'. Accessed: 17-06-2022.

URL: *https://destinationcenter.org/wp-content/uploads/2021/10/WTTC-Economic-Impact-Report-2020_vanuatu_A4-24pp-1.pdf*

Yan, L., Gao, B. W. & Zhang, M. (2017), 'A mathematical model for tourism potential assessment', *Tourism Management* **63**, 355–365.

Yang, Y. & Fik, T. (2014), 'Spatial effects in regional tourism growth', *Annals of Tourism Research* **46**, 144–162.

Yang, Y. & Wong, K. K. (2012), 'A spatial econometric approach to model spillover effects in tourism flows', *Journal of Travel Research* **51**(6), 768–778.

Ľudmila Jánošíková (n.d.), 'Matica vzdialeností'. [Data file] Accessed: 31-07-2022.

URL: *http://frdsa.fri.uniza.sk/~janosik/*

Štátna Ochrana Prírody Slovenskej Republiky (2020), 'Zoznam mchú k

31.12.2020'. [Data file] Accessed: 16-02-2021.

URL: *http://www.sopsr.sk/web/?cl=114*

Appendix A Maps

The author created the following maps for the most important variables using Datawrapper on the website: `https://www.datawrapper.de/`. The values for Bratislava districts were not included; all four districts of Košice are presented as one and therefore have the same color.

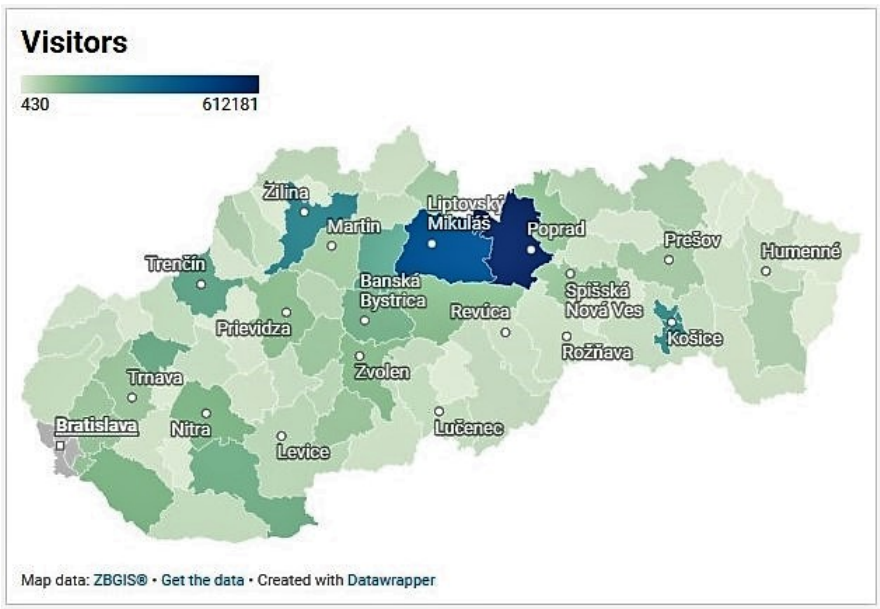

Figure 4: Map of the number of visitors

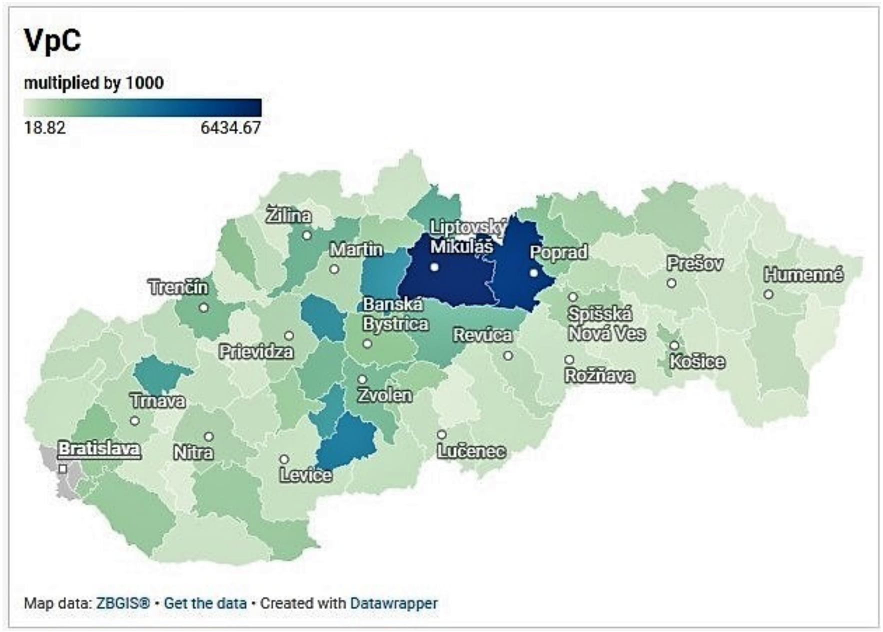

Figure 5: Map of the number of visitors per capita (multiplied by 1000)

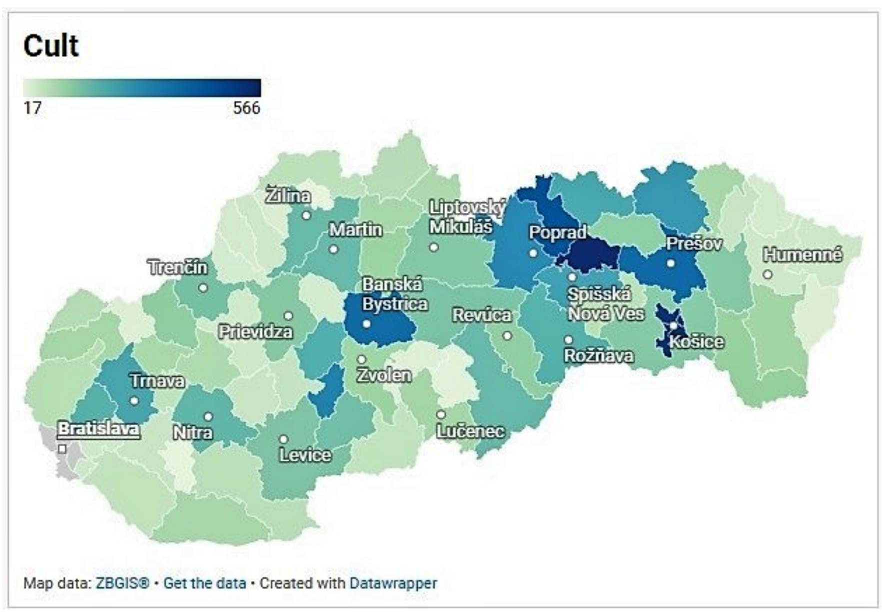

Figure 6: Map of the number of national cultural monuments

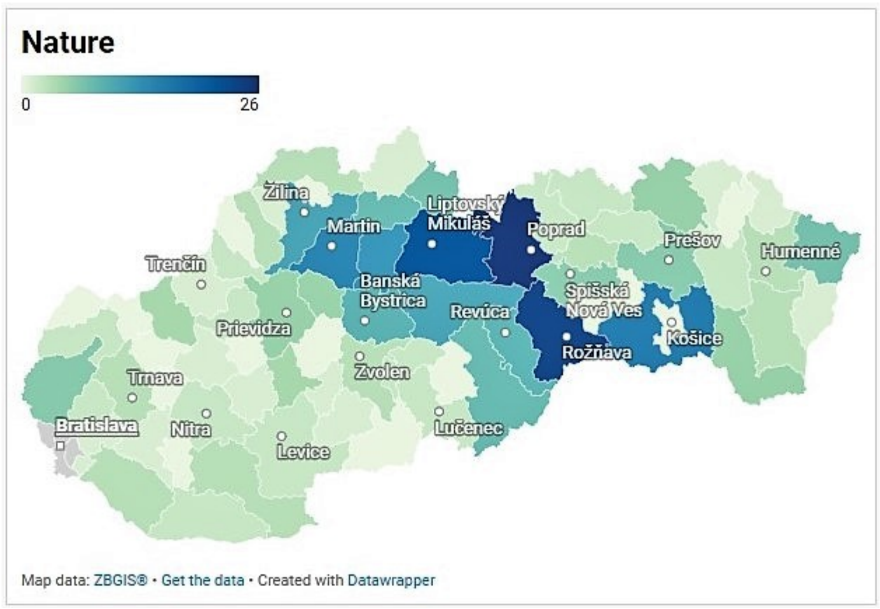

Figure 7: Map of the number of national natural monuments and reserves

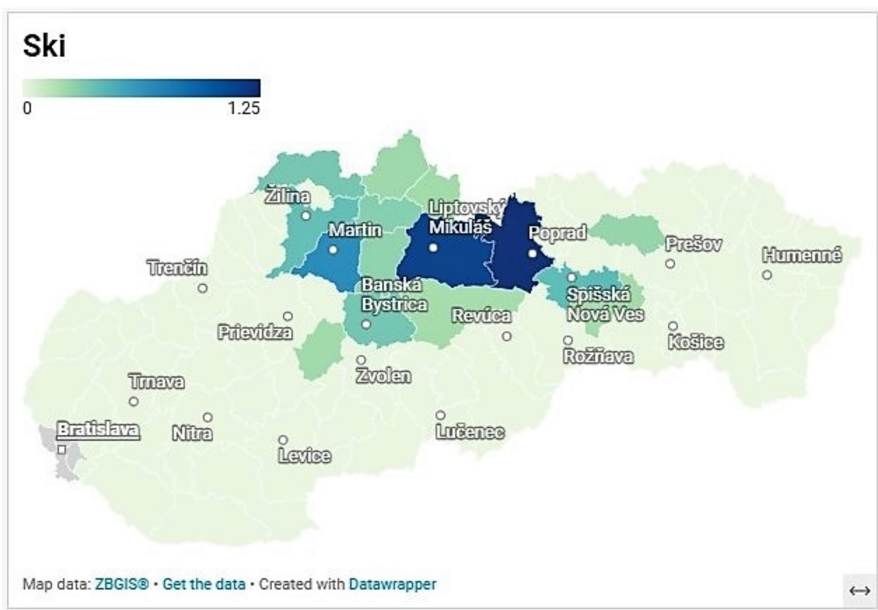

Figure 8: Map of the sum of indices assigned to every ski resort (with an index of at least 0.2)

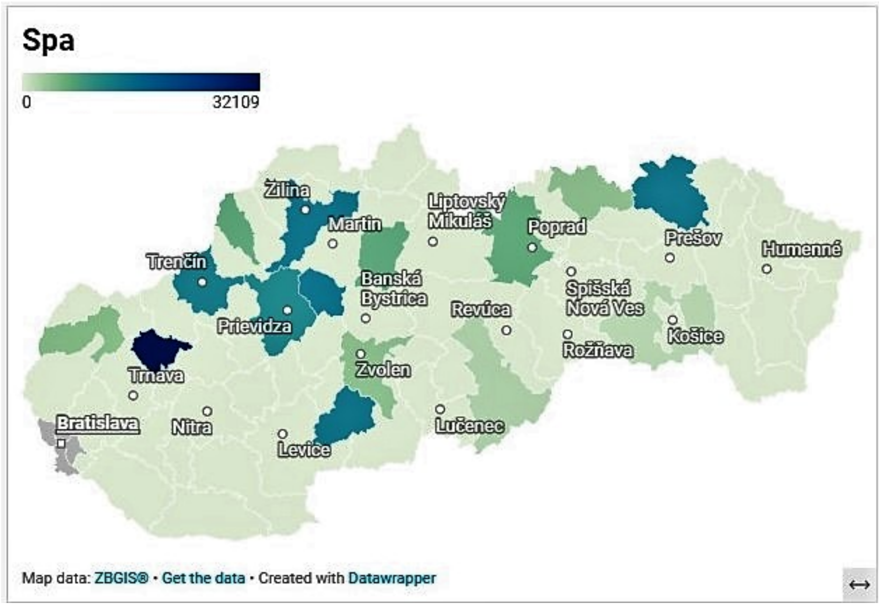

Figure 9: Map of the sum of sales of spa resorts per district

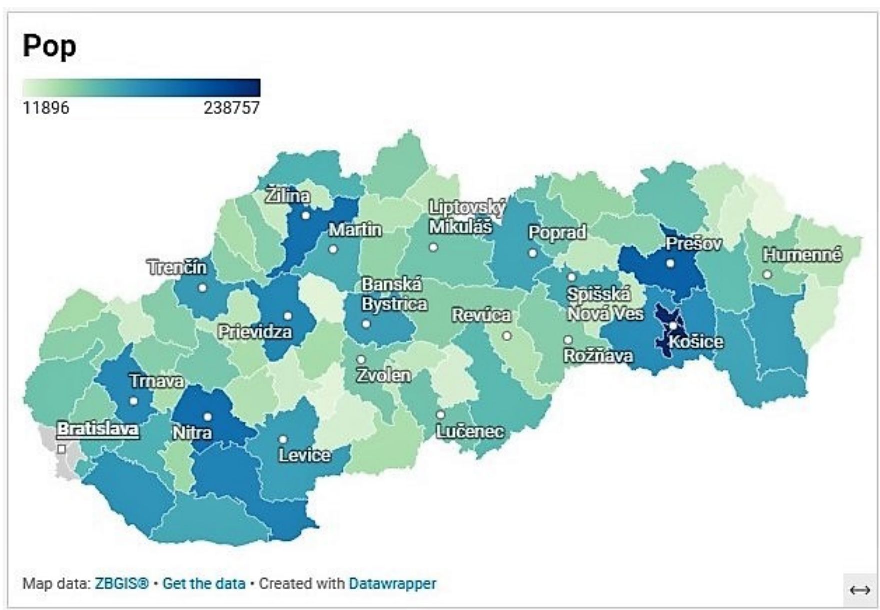

Figure 10: Map of the number of inhabitants

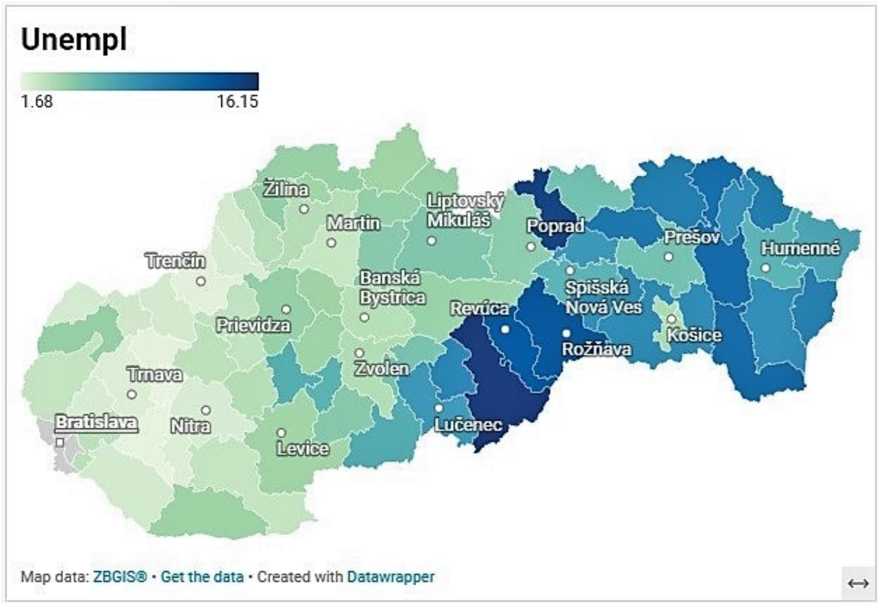

Figure 11: Map of the rate of unemployment

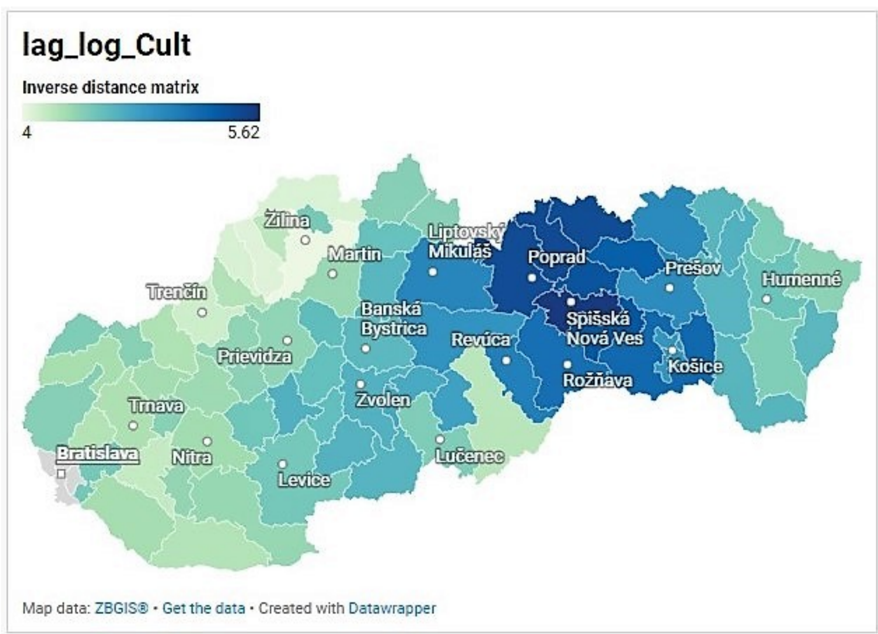

Figure 12: Map of the spatial lags of variable log_Cult (Inverse-dist. matrix)

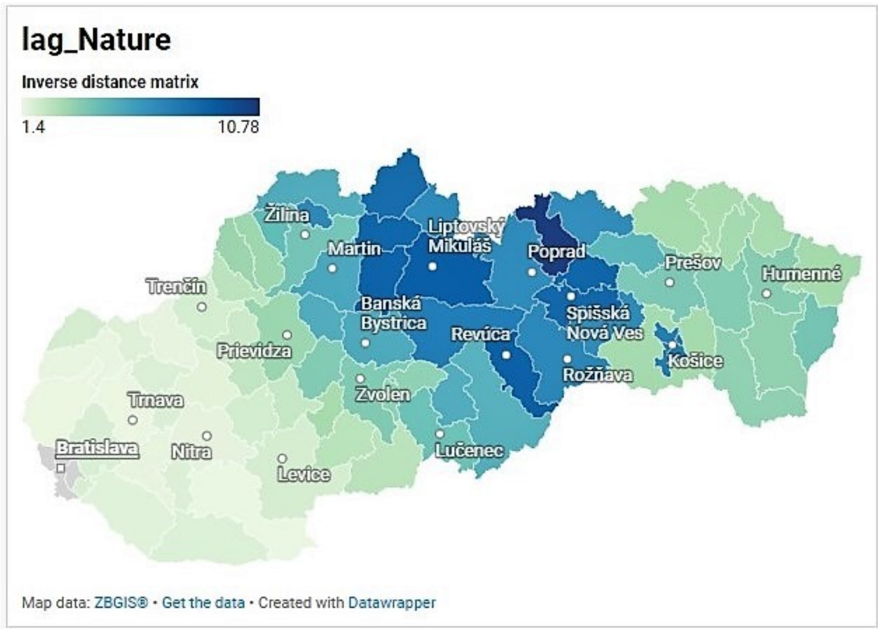

Figure 13: Map of the spatial lags of variable Nature (Inverse-dist. matrix)

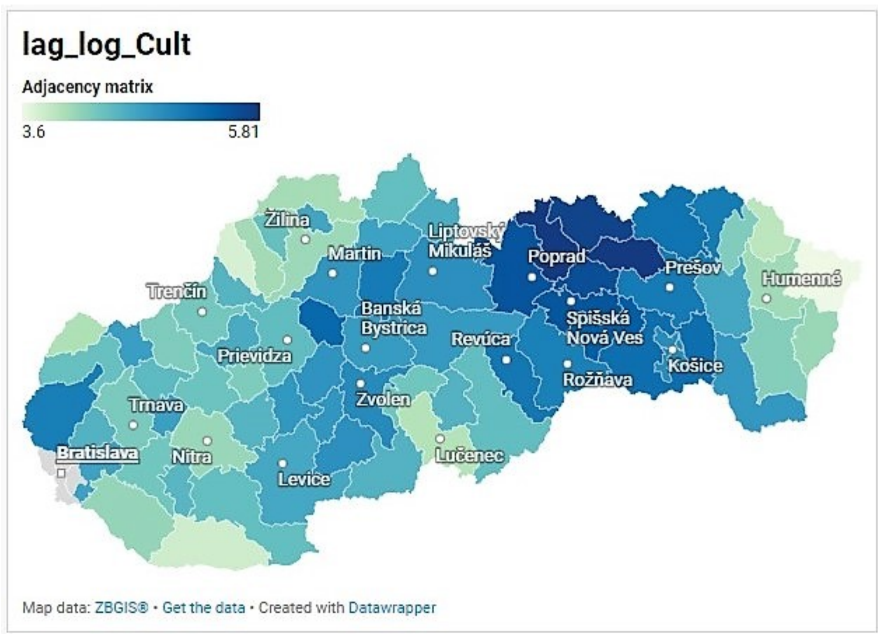

Figure 14: Map of the spatial lags of variable log_Cult (Adjacency matrix)

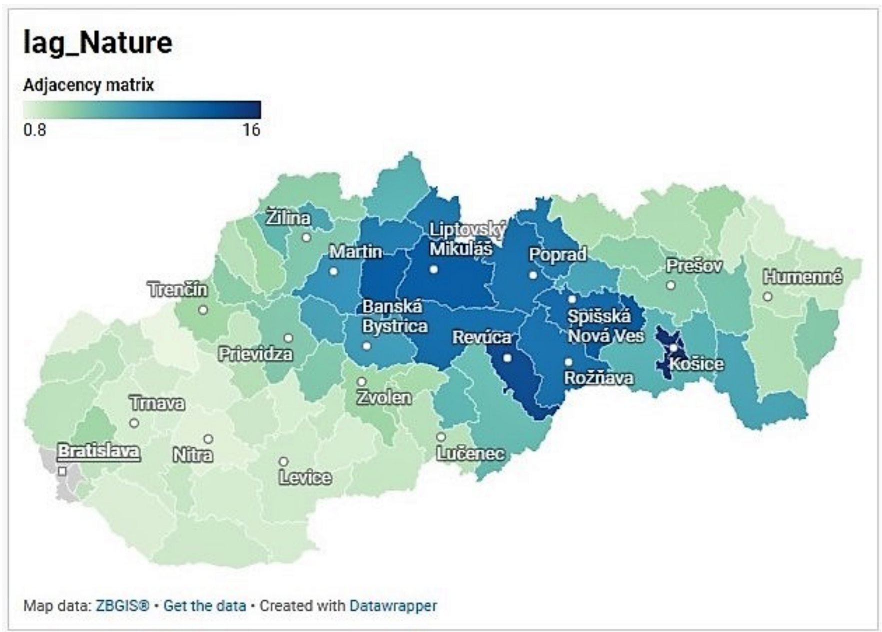

Figure 15: Map of the spatial lags of variable Nature (Adjacency matrix)