IRSTI 71.37.75 https://doi.org/10.26577/JGEM202579410

L.S. Spankulova[1], Y. Yerbolat[2], Y.R. Dauletkhanova[1*]

[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[2]Abai Kazakh National Pedagogical University, Almaty, Kazakhstan
*e-mail: daulethanova_e@mail.ru

ASSESSING TOURISM AND REGIONAL ECONOMIC INEQUALITY IN CHINESE CITIES: SPATIAL ANALYSIS BASED ON POI BIG DATA AND MACHINE LEARNING

This paper presents an overall examination of the interlinkage relationships concerning regional policy, tourism, and economic inequality through the Chinese case. The research objectives are to evaluate the causal effect of the Western Development Strategy (WDS) and provide information regarding the spatial distribution of tourism facilities at the level of Beijing city through the application of contemporary data and advanced methodologies. To achieve the objectives of this study, there are two interconnected empirical research components. The first research component provides a compilation of findings from existing spatial Regression Discontinuity Design (RDD) studies concerning the impact of the Western Development Strategy (WDS). In addition, descriptive statistics from the years 2000 to 2020 are used. The findings of the mentioned researches confirm the existence of a positive effect of the WDS regarding the tourism sector across the targeted regions. This effect is demonstrated through the relative enhancement of the gross regional product's tourism revenue part at a level of approximately 5.9%–6.7 %. The results of the mechanism approach confirm the indirect support of the WDS regarding the tourism sector through the enhancement of the relevant investments in the sector's infrastructure and the extension of the Tax Incentive Schemes. The second research component investigates the spatial distribution of tourism and leisure facilities in the primary city districts of the city of Beijing through the application of POI big data information concerning the relevant sector. Additionally, the findings of the research will be used concerning the information of the relevant sector's POI big data. Machine learning algorithms and decision trees will be employed for the identification of the best locations suitable for the allocation of tourism facilities. The accuracy level of the model achieves the remarkable figure of approximately 83.5 %. The four basic factors that affect the spatial distribution of tourism facilities are hotel density, vicinity to shopping malls, transport accessibility levels, and the relevant sector's POI big data information regarding the city's relative population. The results can contribute to the development of an empirical standard regarding the RDD method in the field of tourism economics and the application of POI big data information concerning AI through the enhancement of the effect of regional policy concerning the mitigation of regional inequality levels.

Keywords: urban tourism, inequality, POI data, machine learning, China.

Л.С. Спанкулова[1], Е. Ерболат[2], Е.Р. Дәулетханова[1*]

[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
[2]Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы, Қазақстан
*e-mail: daulethanova_e@mail.ru

Қытай қалаларында туризм мен аймақтық экономикалық теңсіздікті бағалау: POI үлкен деректері мен машиналық оқытуға негізделген кеңістіктік талдау

Бұл мақалада Қытай мысалында аймақтық саясат, туризм және экономикалық теңсіздік арасындағы өзара байланыстар кешенді түрде талданады. Зерттеудің мақсаты – Батысты дамыту стратегиясының (Western Development Strategy – WDS) туризмге әсерін себеп-салдарлық тұрғыдан бағалау және Бейжің қаласы деңгейінде туризм инфрақұрылымының кеңістіктік үлгілерін заманауи деректер мен әдістер арқылы сипаттау. Осы мақсатта зерттеу екі өзара толықтырушы эмпирикалық бағытты қамтиды. Бірінші бағытта WDS аясындағы аймақтық саясаттың туризм дамуына тигізетін әсері кеңістіктік регрессиялық дисконтинуум дизайны (spatial Regression Discontinuity Design – RDD) әдісімен бағаланды. Бұл тәсіл географиялық шекара бойында орналасқан қалаларды «емделуші» (WDS аясында) және «бақылау» (WDS сыртында) топтарға бөліп салыстыру арқылы саясаттың таза әсерін анықтауға мүмкіндік береді. Эмпирикалық нәтижелер WDS-тің туризм секторының дамуына оң және статистикалық тұрғыда мәнді әсері бар екенін көрсетті: туризм кірісінің жалпы өңірлік өнімдегі үлесі шамамен 5,9–6,7 %

артқан, ал әсер 2013 жылға дейін тұрақты сақталған. Механизмдік талдау WDS-тің туризмді инфрақұрылымға инвестицияларды ұлғайту және салықтық ынталандырулар жүйесін кеңейту арқылы жанама қолдайтынын көрсетті. Екінші бағытта Бейжің қаласының негізгі қалалық аудандарындағы туризм және демалыс нысандарының кеңістіктік үлгілері POI (Point of Interest) үлкен деректері, демографиялық көрсеткіштер және көлік инфрақұрылымы деректері негізінде зерттелді. Машиналық оқыту әдістері, соның ішінде шешім ағашы моделі, туристік нысандарды орналастырудың оңтайлы аймақтарын анықтау үшін қолданылды; модельдің болжамдық дәлдігі 83,5 %-ды құрады. Туризм инфрақұрылымының шоғырлануына қонақүйлердің тығыздығы, сауда орталықтарының жақындығы, көлік қолжетімділігі және халық тығыздығы басты факторлар ретінде айқындалды. Зерттеу нәтижелері туризм экономикасында RDD әдісін қолдану арқылы себеп-салдарлық талдаудың эмпирикалық стандартына үлес қосып, POI үлкен деректері мен машиналық оқытуды біріктірудің аймақтық теңсіздікті азайту және қалалық жоспарлауды оңтайландыру үшін практикалық мүмкіндіктерін көрсетеді.

Түйін сөздер: қалалық туризм, теңсіздік, POI деректері, машиналық оқыту, Қытай.

Л.С. Спанкулова[1], Е. Ерболат[2], Е.Р. Дәулетханова[1]*

[1]Казахский национальный университет имени аль-Фараби, Алматы, Казахстан
[2]Казахский национальный педагогический университет имени Абая, Алматы, Казахстан
*e-mail: daulethanova_e@mail.ru

Оценка туризма и регионального экономического неравенства в городах Китая: пространственный анализ на основе POI-больших данных и методов машинного обучения

В статье комплексно анализируются взаимосвязи между региональной политикой, туризмом и экономическим неравенством на примере Китая. Цель исследования – причинно-следственная оценка влияния стратегии развития Запада (Western Development Strategy – WDS) на туризм и описание пространственных моделей туристской инфраструктуры на уровне города Пекин с использованием современных данных и методов. Для достижения этой цели работа включает два взаимодополняющих эмпирических направления. В первом направлении влияние региональной политики в рамках WDS на развитие туризма оценивается с помощью метода пространственного регрессионного дисконтинуум-дизайна (spatial Regression Discontinuity Design – RDD). Этот подход позволяет выявить чистый эффект политики за счёт сравнения городов, расположенных вдоль географической границы, отнесённых к «лечебной» (внутри WDS) и «контрольной» (вне WDS) группам. Эмпирические результаты показывают положительное и статистически значимое влияние WDS на развитие туристского сектора: доля доходов от туризма в валовом региональном продукте увеличивается примерно на 5,9–6,7 %, при этом эффект сохраняется до 2013 года. Механизмный анализ демонстрирует, что WDS косвенно поддерживает туризм за счёт увеличения инвестиций в инфраструктуру и расширения системы налоговых стимулов. Во втором направлении исследуются пространственные паттерны размещения туристских и рекреационных объектов в основных городских районах Пекина на основе данных POI (Point of Interest), демографических показателей и информации о транспортной инфраструктуре. Методы машинного обучения, включая модель дерева решений, использованы для выявления оптимальных зон размещения туристских объектов; прогностическая точность модели составляет 83,5 %. В качестве ключевых факторов концентрации туристской инфраструктуры выявлены плотность гостиниц, близость торговых центров, транспортная доступность и плотность населения. Результаты исследования вносят вклад в формирование эмпирического стандарта причинно-следственного анализа в экономике туризма с применением RDD и демонстрируют практический потенциал сочетания больших данных POI и машинного обучения для снижения регионального неравенства, и оптимизации городского планирования.

Ключевые слова: городской туризм, неравенство, данные POI, машинное обучение, Китай.

Introduction

Urban tourism and the provision of leisure facilities in the city can be considered a crucial part of modern-day city development and regional economic policy. This development helps the efficient coordination of the city's infrastructure network and increases its tourism value, which also leads to the improvement of the city's residents' quality of life and helps develop the city's economy. The city's tourism also promotes its global image and has a multiplier effect in the labor market and the service sector (Loh, W.Y., 2011: 14).

In the past few years, the city has grown to be one of the preeminent tourism destinations in the nation. The city's tourism industry has been projected as one of the prime factors influencing the growth of the economy due to the inflow of visitors from across the world. The above-mentioned points give sufficient reasoning to study models of tourism development in the following ways (Figure 1).

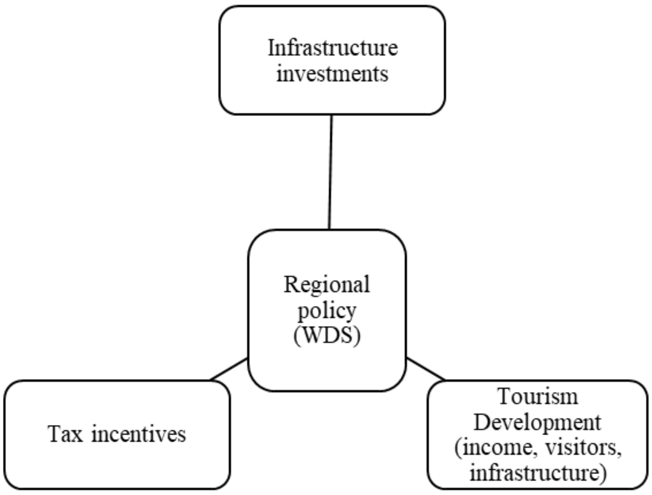

Figure 1 – Concrete models of tourism development
Note – compiled by the author

The above-mentioned factors currently require drastic research through various models of studying the opinions of the general public through various surveys and observations. However, in recent years, the increasing usage of big-data analytics to promote decision-making in the field of geographical locations has raised popularity due to its capability to provide information regarding the location through various models of POI data points, geographical information of various social sites, and location information of various mobile network operators (Wu, S., 2025: e0298056).

In addition, the regional development policy of the PRC has also had the same effect on the geographic pattern of tourism development. Since the beginning of the twenty-first century, the Chinese government has been promoting the Western Development Strategy (WDS). The aim of the WDS has been to eliminate the development lag of the western part of the PRC and to address the inequality of society as a whole. The WDS has been identified as the longest regional policy of the post-reform era. In the context of the WDS, the target has been to improve the economic prospects of the twelve provinces and the development prospects of the five autonomous regions. The first stage of the research focused on analyzing the trends in tourism revenue, detailed in the following table (Table 1).

Table 1 – Tourism revenues in China, 2000–2020 (in billion USD). The table includes data to illustrate the trends and key indicators of the tourism sector.

Year	Domestic tourism revenue (bn USD)	Inbound tourism revenue (bn USD)	Share of the western region (%)
2000	317 bn USD	100 bn USD	10 %
2010	1 257 bn USD	228 bn USD	18 %
2020	3 995 bn USD	440 bn USD	25 %
Note – compiled by the author			

As shown in the table, tourism revenues from domestic and inbound flows in China demonstrated a steady growth trajectory between 2000 and 2020. In 2000, domestic tourism generated 317 billion yuan and inbound tourism 100 billion yuan, whereas by 2020 these figures had increased to 3,995 billion yuan and 440 billion yuan respectively. Thus, revenues from inbound tourism grew by approximately a factor of 4.4. From a regional perspective, the share of tourism income attributable to the western regions of China also rose markedly – from 10% in 2000 to 25% in 2020. This shift underscores the important role of tourism in stimulating economic activity in the western and central provinces of the country and in mitigating interregional income disparities (Fan, C. C., 2008: 18).

In addition, the expansion of tourism has had a positive effect on the national labour market, particularly by contributing to the creation of new jobs in the service sector (Table 2).

Table 2 – Changes in employment in tourism and the service sector in China, 2000–2020 (in thousands of employees). The table includes data to illustrate the trends and key indicators in employment within the tourism and service sectors.

Industry	2000 year (number of people)	2020 year (number of people)	Change (%)
Tourism and services	3.1 million people	8.9 million people	+187%
Industry and construction	10.5 million people	14.8 million people	+41%
Note – compiled by the author			

The findings indicate that between 2000 and 2020, the number of people employed in tourism and related service sectors in China increased substantially. Whereas in 2000 these sectors employed 3.1 million people, by 2020 this figure had risen to 8.9 million, representing an increase of 187%. For comparison, employment in industry and construction grew by only 41% over the same period. These data demonstrate that tourism has become an important economic factor in diversifying the national labour market and generating new jobs (Gao, S., 2017: 446–467). Furthermore, the development of the tourism industry and increased investment in regional infrastructure have contributed to a marked reduction in poverty levels in the western regions (Table 3).

Table 3 – Dynamics of poverty levels in Western China, 2000–2020

Year	Share of the population living below the poverty line %
2000	20.3%
2010	10.5%
2020	0.6%
Note – compiled by the author	

According to Table 3 above, the rate of poverty in the western regions of China has dropped from 20.3% to 0.6% from the year 2000 to the year 2020. This makes it clear that the development of tourism and the economic policy of the region has been quite effective. The development of the tourism sector through the development of the sector's infrastructure has helped in improving the income levels of the people and has reduced the levels of poverty by a factor of 33 (Qin, X., 2023: 359-385).

The existence of rich natural, historical, and cultural resources in the western region has facilitated the positioning of the tourism sector as the "backbone" of the regional economy. According to statistical information, from 2002 to 2010, the revenue of the tourism sector of Western China recorded a dramatic growth from 18.36 billion US dollars to 103.7 billion US dollars, while the number of tourists also jumped from 304.6 million to over 1 billion. However, there has been insufficient research carried out from the scientific community about whether this growth can be accredited to the direct results of the application of the WDS policy or the overall growth of this sector at the national level (Qian et al., 2021: 101552).

In this context, the objectives of the current study are to find the causal effect of regional policy on the development of tourism and to develop efficient ways of planning city tourism based on spatial information. In general, the study will address the following objectives:

- to assess the effect of WDS on the tourism sector within a causal framework through a spatial Regression Discontinuity Design (spatial RDD).

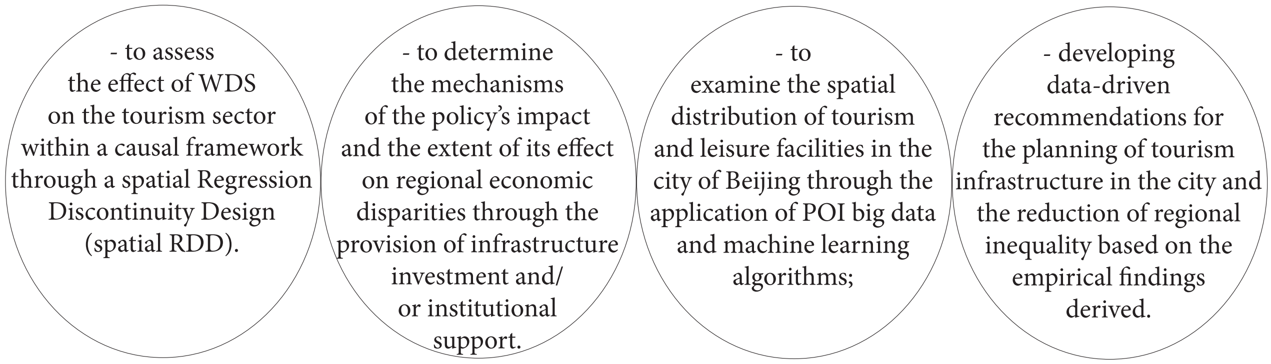

Figure 2 – Conceptual framework of the study
Note – compiled by the author

- to determine the mechanisms of the policy's impact and the extent of its effect on regional economic disparities through the provision of infrastructure investment and/or institutional support.

- to examine the spatial distribution of tourism and leisure facilities in the city of Beijing through the application of POI big data and machine learning algorithms;

- developing data-driven recommendations for the planning of tourism infrastructure in the city and the reduction of regional inequality based on the empirical findings derived.

The scientific significance of the problem: the relevance of the problem consists in the fact that the tourism sector is a complex system that relies on the multisectoral economy. The development of the latter occurs to a large extent depending on the type of regional policy, the level of the state's infrastructure development, as well as the quality of spatial planning (Li, X., 2020: 100608). The previous researches mainly estimated the impact of policy through the comparative method, which has the disadvantage of endogeneity. To overcome this drawback, the current research intends to apply the method of the spatial Regression Discontinuity Design (spatial RDD) to measure the exact causal impact of the policy. The study will be based on the framework of spatial RDD & POI big data & CART algorithms.

The results provide new insights into the relationships between regional inequality, the development of transportation infrastructure, and tourism. By integrating the results from the RDD estimation method with those obtained through machine learning algorithms applied to tourism economics, the study not only establishes a novel research framework but also produces detailed, data-driven findings that can support the National Bureau of Statistics of China in creating more accurate and evidence-based datasets for evaluating regional policy effectiveness (National Bureau of Statistics of China, 2010). Thus, integrating the methodology of spatial regression discontinuity design (RDD) with POI big data analytics and machine learning algorithms creates not only a new research toolkit, but also a practical mechanism for improving the quality of official statistics. The data obtained in the course of the study make it possible to move from stating the facts of growth in tourism revenues, for example, in the western regions of China, to an evidence-based assessment of the contribution of a specific policy – the Western Development Strategy (WDS)- to this growth. The 'pure' causal effect of 5.9–6.7% identified using RDD provides statistical agencies, such as the National Bureau of Statistics of China, with a benchmark quantitative indicator of the effectiveness of such large-scale regional programmes.

In addition, the model developed for Beijing based on POI data and the CART algorithm demonstrates how big data and machine learning can serve as the basis for creating detailed spatially referenced datasets. Instead of averaged indicators for a city or district, this methodology allows the generation of data with high geographical resolution (at the 1x1 km level), directly linking the location of tourist infrastructure with population density, transport accessibility and the availability of related services (hotels, shopping centres). This enables the National Bureau of Statistics to generate not just reporting data sets, but also analytical and predictive data sets. Such data can be used to accurately assess the impact of infrastructure projects that have already been implemented, as well as to model the potential effect of planned investments, which is extremely important for justifying budget allocations and strategic planning with a view to reducing regional inequality. In fact, the study offers a roadmap for enriching traditional statistical accounting with evidence-based, causal and spatially detailed insights.

The target of this research study will be regional development policy and the tourism sector in the PRC, which will also encompass the regions affected by the Western Development Strategy and the large city aggregates.

The research subject involves the role of regional policy in developing the tourism sector and reducing regional economic disparity. In addition to this, the spatial distribution of tourism and leisure facilities in the core districts of the city of Beijing also forms the research subject.

The main hypothesis of the study is that the Western Development Strategy leads to the development of the tourism sector due to the enhancement of infrastructure investment and the provision of tax incentives, which ultimately helps in the mitigation of regional economic disparities, and the spatial distribution of tourism and leisure facilities can be predicted through the use of POI big data, the density of the population, and transport accessibility.

It is important to note that while the first research component focuses on the Western Development Strategy (WDS) and its impact on tourism in western regions, the second component considers Beijing as a separate case study. Beijing, as the capital and a major tourist centre in eastern China, was chosen not as a WDS target but as a case study of a developed metropolis where advanced spatial analysis using big POI data and machine learning can be applied. This two-tiered approach allows us to: 1) assess the causal relationships of regional policy (WDS) at the macro level and 2) demonstrate a modern methodology for optimising tourism infrastructure planning at the city level, which can subsequently be applied to cities within WDS regions.

Literature review

The interplay of tourism and regional economic development has, during the last decade, emerged as one of the main research trends in the field of regional science and economic geography. A number of studies have revealed the positive effect of tourism as a factor of economic growth and regional income distribution (Croes & Vanegas, 2008; Yang et al., 2021). On the other hand, researchers mentioned above also identify that irregular development of tourism can contribute to the aggravation of socio-economic differences between regions (Liu, M., 2020; Fang et al., 2021). This can be especially topical concerning the situation of large nations like China.

In an effort to rectify regional disparity issues, the Chinese government introduced the Western Development Strategy (WDS). This has been regarded as one of the significant national initiatives concerning the geographic redistribution of economic and social development (Lu, 2011 & 2013; Grewal, 2011; Zheng et al., 2022). The main purpose of the WDS has been to unlock the development possibilities of the less-developed western regions through improved transport and tourism infrastructures, attractiveness of investment opportunities, and increased government subsidies (Chen et al., 2013). Although it has been found in previous studies that the WDS has positively affected regional economic development and investment (Démurger et al., 2002 & Wang & Wei, 2020), its effect remains unestablished from a causal point of view in the tourism sector.

In recent years, spatial data and various types of big data have been used widely to evaluate the development of tourism. In particular, the available POI (Points of Interest) data allows the functional structure of cities, the level of tourism attractiveness, and the intensity of the service sector's activity to be described accurately (Lu et al., 2020; Xu, 2024; Yi, D., Yang, J., Liu, J., Liu, Y., & Zhang, J. 2019). In this study, the POI data and the ML approach will also be used to study the geographical distribution of tourism and entertainment facilities in the main city districts of the Chinese capital of Beijing. The fact that this information has geographic coordinates at particular locations allows micro-level research of the economic and tourism processes in each of the regions studied. An important trend in this field has recently become the development of new sources of information of the type 'nightlight images,' which can be used alternatively as a new indicator of regional economic activity (Gibson, 2021).

Although classical econometric methods, like Difference-in-Differences and Fixed Effects models, remain widely used to study the relationship between tourism and regional policy, they do not consider spatial interdependencies and border effects properly (Anselin, 2010). During the last few years, in order to overcome the above limitations of previous methodologies, the Regression Discontinuity Design method incorporating spatial information (spatial RDD) has been employed rather frequently (Keele & Titiunik, 2015; Dell, 2010). Using spatial

RDD, it becomes possible to isolate the causal effect of a policy through the identification of cities along the border of the policy effect and identify the effect of geographical localization accurately. This method sets a new standard in the evaluation of the effectiveness of regional development and tourism policy (Dong et al., 2019).

In the study of the relationship between the development of tourism and regional inequality, the role of machine learning algorithms has also gained increasing significance. Using algorithms involving Random Forest, XGBoost, and GWR (Geographically Weighted Regression) allows researchers to address the complications of multiple factors in the development of tourism and identify spatial patterns (Du et al., 2023). By combining POI information and information from social media sites, the above algorithms can accurately predict the density of tourism development in cities and intercity differences and the impact of policy intervention (Xu, 2022).

In summing up the existing modern research literature, the complexity of the data level involved in analyzing the interaction of tourism and regional policy problems and the significance of spatial causal analytics are revealed. The WDS and other regional initiatives represent direct policy experiments at which the problem of regional inequality in the development of tourism in the PRC can be mitigated. To correctly interpret the effect of the described regional initiatives in concrete empirical research will require novel research designs involving spatial data (data of POI, nighttime lights data, and big social media data), together with the application of machine learning methodologies.

Therefore, although there has been relevant research concerning the interrelated topics of tourism, regional policy, and regional economic inequality from diverse perspectives, there has been insufficient research concerning the effect of the Western Development Strategy upon the tourism sector from the viewpoint of spatial RDD in the context of a causal model and the spatial distribution of the tourism infrastructure of cities from the intersection of POI big data and the application of ML algorithms. This paper aims to fill this research gap.

Materials and methods

The research examines the relationships existing between regional policy, tourism, and city infrastructure at two interrelated scales: the national and regional level as framed within the Western Development Strategy (WDS), and the intra-urban level in the central districts of the city of Beijing.

The research question being asked is: What is the impact of regional policy on the development of tourism and regional disparities, and to what extent can the distribution of tourism and leisure facilities in a large city be predicted through the usage of POI big data and the density of the population? The hypothesis of this research is that the WDS has a positive impact on the development of tourism due to its supporting role in the development of the region's infrastructure and thus leads to the reduction of regional disparities, and at the same time the distribution of tourism and leisure facilities can be predicted.

From the methodological point of view, the research combines three elements: (1) a descriptive analysis of national and regional statistics, (2) the synthesis of existing spatial RDD research concerning the WDS, and (3) the micro-level model of POI big data and the CART algorithm used in the context of Beijing.

Study areas and data

The macro-level analysis covers all regions affected by the OBD. Beijing was deliberately chosen for the micro-level study, even though it is not a target city for the OBD. As a global metropolis with a developed tourism sector and high data availability (POI, infrastructure, population), Beijing serves as an ideal testing ground for the proposed POI methodology and machine learning. The ideas and models obtained in Beijing are designed to be used for analysing and planning tourism infrastructure in other cities, including in western China, within the framework of the WDS.

In the macro approach, the research area extends to the land of China with the emphasis being on the regions defined in the WDS. The relevant data concerning earnings from domestic and incoming tourism, the number of people employed in tourism and the services sector, as well as the level of relative poverty from the years 2000–2020 were collected from the government publication of the National Statistical Office of China (China Statistical Yearbook). The purpose of the data collection was to develop narrative statistical tables concerning the earnings and the fall of relative poverty levels of the regions of the west.

A micro-level study will target the main urban districts of Beijing because this region constitutes one of the most advanced and complex environments of its type in the whole of China. Three main sources of data were used:

- POI data from the AutoNavi (Amap) platform, containing the coordinates of 566,932 facilities as of October 2023.

- population data from the WorldPop project in the year 2020, with a spatial resolution of 100 meters;

- urban infrastructure and transport information from the official statistical reports as well as OpenStreetMap.

The study area was partitioned into a regular grid of squares of size 1 x 1 km. This resulted in the formation of 1548 valid grid squares. The number of facilities and the existence of facilities for each of the grid squares were measured through a series of points of interest (POI) classifications based on the functional classification of the city.

Macro-level component: WDS, tourism and regional inequality

The macro-level research investigates the causal effect of the Western Development Strategy (WDS) of China on the development of the tourism industry.

In previous researches, the hypothesis that there is a larger contribution of tourism to the GRP in regions located inside the WDS than in their neighboring regions has been tested. For the current article, instead of re-estimating the previous models, the results of the previous researches will be used.

A large number of studies used spatial Regression Discontinuity Design (spatial RDD) to compare regions inside the WDS framework to those regions which lie outside the framework. This method takes advantage of the policy boundary and controls the initial similarity of the territories to better identify the causal effect of regional policy on the growth of tourism income.

In our research, the above-mentioned findings at the macro-level of research are combined to outline the general trends of tourism development during the WDS. Proceeding from this general outline of the trends of tourism development during the WDS, the micro-level research gives a particular emphasis to the example of the city of Beijing and utilizes the big data of the points of interest and the possibilities of machine learning algorithms applied to this data to identify the spatial distribution of the tourism infrastructure of the city.

Micro-level component: POI big data and CART model for Beijing

The case study region selected was the city of Beijing, which is the political and cultural capital of the People's Republic of China. This city reflects one of the regions in the country which has been heavily urbanized, whereby the development of tourism and recreation has also been spatially concentrated. The region of study will be the urban region of the city of Beijing.

The study area was partitioned into a uniform grid of cells of size 1 × 1 km. A total of 1,548 grid cells were defined, enabling the distribution of facilities in the city to be examined in a relatively homogeneous fashion. Indicators were summed together at the level of each cell regarding the factors of population, tourism facilities, and public service facilities.

The data sources and characteristics of the data are:

- POI data from the AutoNavi (Amap) platform – 566,932 spatial points as of October 2020, representing the geographical locations of particular facilities in various sectors.

- population data from the WorldPop project (2020) – raster data of population density with a spatial resolution of 100m.

All points of interest were coded under 14 primary groups, which included transport infrastructure, educational institutions, healthcare facilities, public organizations, retail, culture, household services, facilities for leisure and tourism, and others. Per grid cell, the existence and non-existence of the above interest facility types were noted through binary coding (0 – absent, 1 – existing).

The research work involved various stages:

1. Spatial grid formation and POI data integration. The count and distribution of POI points were calculated for each of the 1 × 1 km grid cells by aggregating the points within each grid cell.
2. Binary Encoding & Dataset Creation. The grid cells representing regions containing tourism & leisure activities and regions without tourism & leisure activities were coded through binary indicators.
3. Balancing the data and the development of the machine learning model. An equal number of samples were chosen as positive (when the facility existed) and negative (when the facility did not exist).
4. Application of the Classification and Regression Tree (CART) algorithm. This approach was employed in this study to evaluate the effect of each type of infrastructure studied on the location of tourism facilities.
5. Generating a predictive map. Using the resultant model, a total of 629 sites of high potential for the development of tourism facilities were identified.

6. Model validation. The model resulted in predictive accuracy of 83.5%, which shows that the model was reliable.

7. Hypothesis testing. The above steps were intended to test the hypothesis that the spatial pattern of tourism and leisure facilities is statistically significantly related to the density of hotels, shops and service enterprises, transport accessibility, and the density of the population.

The new methodological approach has various benefits over conventional data gathering methodologies:

- the capability of processing mass data in a short period of time;
- a high level of spatial precision (ranging from 100 m to 1 km).
- a reduction in the impact of human subjectivity because of the application of machine learning.

In the context of the study, the pre-processing of the data and descriptive statistics, spatial statistics (Moran's Index, MSA), classical regression models, as well as models based on the CART algorithm from the field of machine learning were used collectively.

Certain limitations can be noted in the study: the incompleteness of the POI coverage and the rate of data update may cause slight distortions in the model results. However, the results achieved allow the development of a model of spatial distribution of tourism facilities in Beijing from a scientific point of view.

In general, the combined method involving spatial analytics methodologies and machine learning has proven to be rather effective at planning the tourism infrastructure of large cities. The combined method involving POI big data and maps of the population density of regions can be used for decision-making in planning the tourism sector and can also be applied elsewhere.

Results and discussion

The results of the study reveal clear spatial regularities in the distribution of tourism and leisure facilities within the core urban area of Beijing. By combining POI data with population-density information and conducting the analysis at the level of a 1 × 1 km grid, it was found that tourism facilities are markedly concentrated in the central parts of the city, while they occur far more sparsely in peripheral zones. This pattern reflects the centre-periphery inequality characteristic of metropolitan structures, as noted by Zhang (2018) and Tan et al. (2020). On the basis of these data, a detailed geographic map of the spatial configuration of tourism and leisure facilities was produced (Li, X., 2021: 1213–1231).

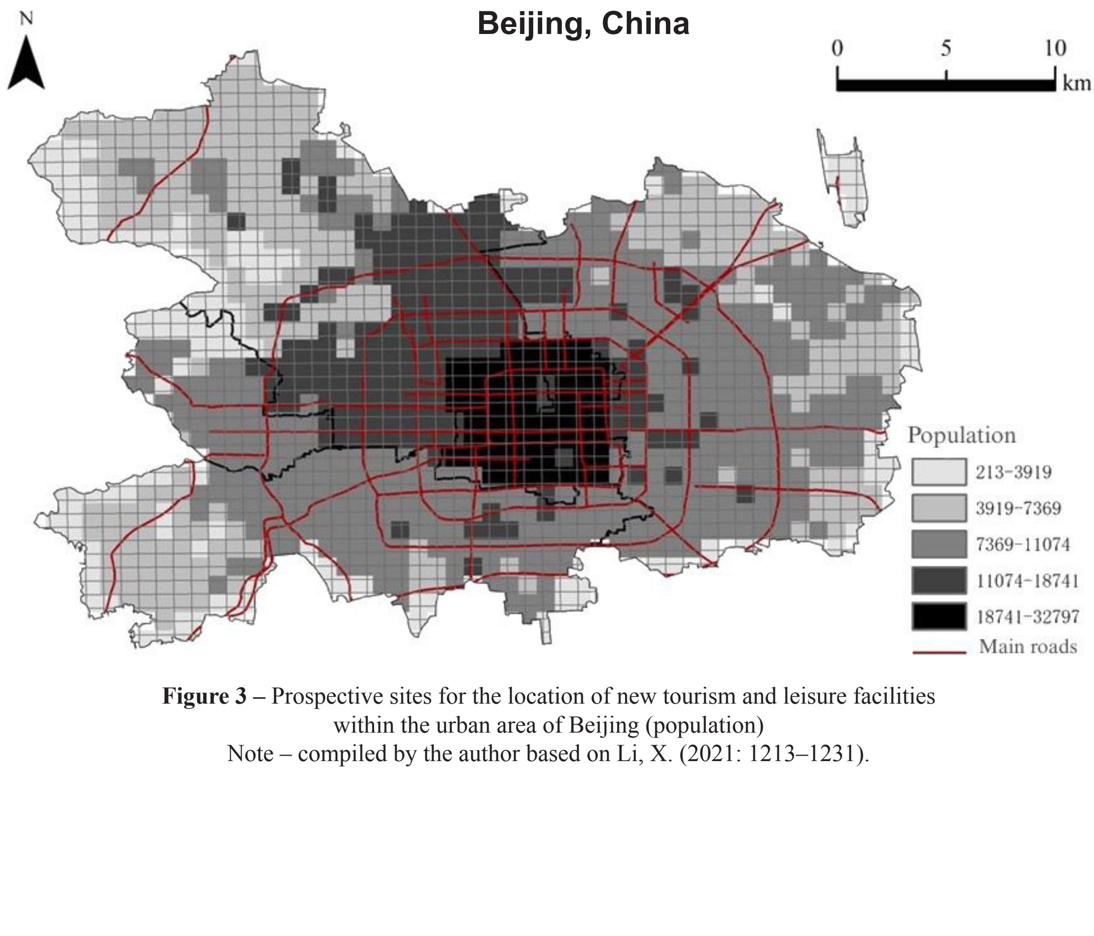

Figure 3 – Prospective sites for the location of new tourism and leisure facilities within the urban area of Beijing (population)
Note – compiled by the author based on Li, X. (2021: 1213–1231).

The POI information of Amap (AutoNavi) contains 566,932 points of interest organized into 14 groups. The data encompasses the city's infrastructure. The main groups are public services, transportation, and household facilities of infrastructure. The entertainment and tourism sector mainly comprises the Dongcheng, Xicheng, and Haidian districts. In the regions mentioned above, the main factor that directly contributes to the development of the tourism sector can be observed as the high density of the region's human population. Using WorldPop information, about 50.2% of the city's population lives in the above regions (He, D., 2021: 12252).

The trends observed are consistent with the findings of Croes & Vanegas (2008) and Yang et al. (2021), which stress the significance of tourism in job creation and income diversification. However, the results of the current study confirm that the extent of this change surpasses national levels and thus reflects the multiplicative impact of the regional policy.

The use of the Classification and Regression Tree (CART) algorithm helped the model accurately predict the possible locations of tourism and leisure facilities in the city. In the training data set, the number of "positive" and "negative" cells was chosen equally from the data set. The algorithm has detected 629 locations of high potential, which are mainly concentrated around the transport network and the city center. The main factors that contribute to the distribution of tourism and leisure facilities in the city are the transport network, residential density, and the provision of public services.

The relevance of hotels and shopping centers relates to the findings of Gao et al. (2017) and Du et al. (2023). They indicate the strong relationship that exists between the density of tourism in cities and the agglomeration of services and shopping centers. Additionally, our model puts a greater significance on the role of population density than the previous model did in the spatial distribution of residential buildings and tourism-leisure facilities.

The model resulted in a predictive accuracy level of 83.5%, which indicates the immense accuracy of the CART model technique used in the study and its ability to accurately predict the tourism potential of the urban area. Moreover, the model accurately described the characteristics of the sizes of the various types of infrastructural components and their contribution to the probabilities of the locations being tourism sites. This indicates a strong relationship that exists between the levels of the city's infrastructural development and the level of tourism.

To illustrate the application of POI data classification in urban studies, Table 4 presents an example from the existing literature that details the classification of objects in Guangzhou (adapted from Li, X., 2025). While the absolute figures are specific to Guangzhou and its unique context, this classification scheme illustrates the type of structural analysis that our study conducts for Beijing. The key point for comparative analysis of cities is not to directly compare the final indicators, but to understand the relative weight and composition of various functional categories (e.g., transport, trade, tourism) in the urban infrastructure.

By comparing the results of the CART model to the case of Beijing, it can be observed that the results of the model regarding the role of the density of hotels and the accessibility of transport in the formation of the distribution of tourism and leisure facilities match the findings of Gao et al. The model results also confirm the role of the density of the population in defining the spatial distribution of the tourism facility distribution: in the high-density central part of the city, the density of the tourism facilities reaches its maximum value. The model results also indicate the existence of high-potential but currently under-explored regions of the peripheral residential districts of the city. Therefore, the results of the model allow planners to identify the regions of priority allocation of the newly emerging tourism and leisure facilities.

As shown in Table 4 (adapted from Li, X., 2025), the distribution of POI categories for Guangzhou shows, for example, that corporate facilities form the dominant group. This shows the rapid growth of businesses and industries in the city. This is followed by transport facilities at 72,247 and food and beverages at 71,080. These three comprise the basic infrastructural elements that contribute to the city's tourism attractivity. The science and education facilities also come in at a high number of 58,500, establishing the city as the intellectual and cultural hub that it is. However, the grouping that relates directly to recreation and tourism has a low number of only 7,366 facilities. However, each one of them has a crucial role in establishing the city's image as a tourist destination (Chen, R., 2020: 637).

Table 4 – Categories of POI facilities in Guangzhou's urban infrastructure and their influence on tourism

Level I category	Level II category	Facility type	Number of POI (units)
Core service facilities	Transport facilities	Airports, railway stations, metro stations, bus terminals, etc.	72 247
	Science, education and culture facilities	Schools, colleges, cultural centres and research institutes	58 500
	Medical facilities	Hospitals, clinics, pharmacies and emergency medical centres	23 278
	Public facilities	Public toilets	15 717
	Household service facilities	Post offices, laundries, beauty salons, offices, etc.	44 040
	Finance and insurance facilities	Banks, ATMs and insurance companies	21 295
Commercial service facilities	Hotels	Hotels, hostels and serviced apartments	4 568
Administrative facilities	Retail	Supermarkets, markets and shopping centres	7 550
	Food and beverage establishments	Restaurants, fast-food outlets and cafés	71 080
	Governmental and public organisations	Government bodies, administrative offices and social institutions	46 016
	Corporate facilities	Companies, cooperatives, agricultural and horticultural organisations	132 613
	Commercial residential buildings	Residential buildings, villas and industrial parks	31 455
Leisure and entertainment facilities	Sports and recreation	Stadiums, cinemas, sanatoriums and playgrounds	31 279
	Tourism and leisure facilities	Tourist attractions, museums, parks, aquariums and resorts	7 366
Total			566 932
Note – compiled by the author			

The findings indicate that the intensity of tourism and leisure facilities' distribution is concentrated in the inner districts, while the number of facilities drops considerably towards the outskirts of the city. The distribution of the facilities can be explained in terms of 'point clustering in the central districts and dispersion in the outer districts,' where the facilities are concentrated in the form of points in the central districts, while the points are dispersed in the outer districts. The optimized model used the machine learning approach to identify suitable sites of 629 locations out of the 1,548 valid grid cells. The classification accuracy of the model showed that 83.5% of the predicted locations actually had the facilities.

The model has also noted particular off-grid regions as having a high level of development potential: Various regions in the outer edge of Haidian District and the Chaoyang District's Western region demonstrated the capability of development even though they are not close to the arterial road network (Deng, T., 2019: 1-16). Out of the 373 grid cells that existed along the 102 streets in the high-density regions, the target locations of high development priority were chosen to be 240. They mostly lie in Dongcheng, Xicheng, Chaoyang, Haidian, FengTai, and Shijingshan districts and others in the peripheral regions.

The predictive analysis shows that the application of POI data and the approach of ML algorithms enables the objective optimization of the spatial distribution of tourism and recreation facilities in the city to a large extent, reducing the level of subjectivity in the decision-making process. The decision model built on the CART algorithm presents explicit recommendations regarding the location of new facilities, which demonstrates its high practical value for the management of the city's infrastructure and tourism. Figure 3 shows the final locations of the suggested sites: in total, 240 new tourism facilities are expected to be built in the high-density regions (marked in green color).

The results of the statistical dynamics conclusively show the impact of tourism on regional economic development and social factors under the WDS. During the years from 2000 to 2020, the growth of the number of people employed in the tourism industry greatly outpaced the average growth of the general economy, supporting the finding that tourism has played a role in developing the labor market through the development of the service sector. The findings also support the views of Croes & Vanegas and Li regarding the impact of tourism on the labor market.

In addition, the fact that the level of poverty has dropped considerably in the WDS regions attests to the indirect impact of tourism in redistributing wealth, as described in the work of Fang and others. On the other hand, the findings in this study also confirm that the economic base and transport infrastructure of regions where the role of tourism has grown have played an important role.

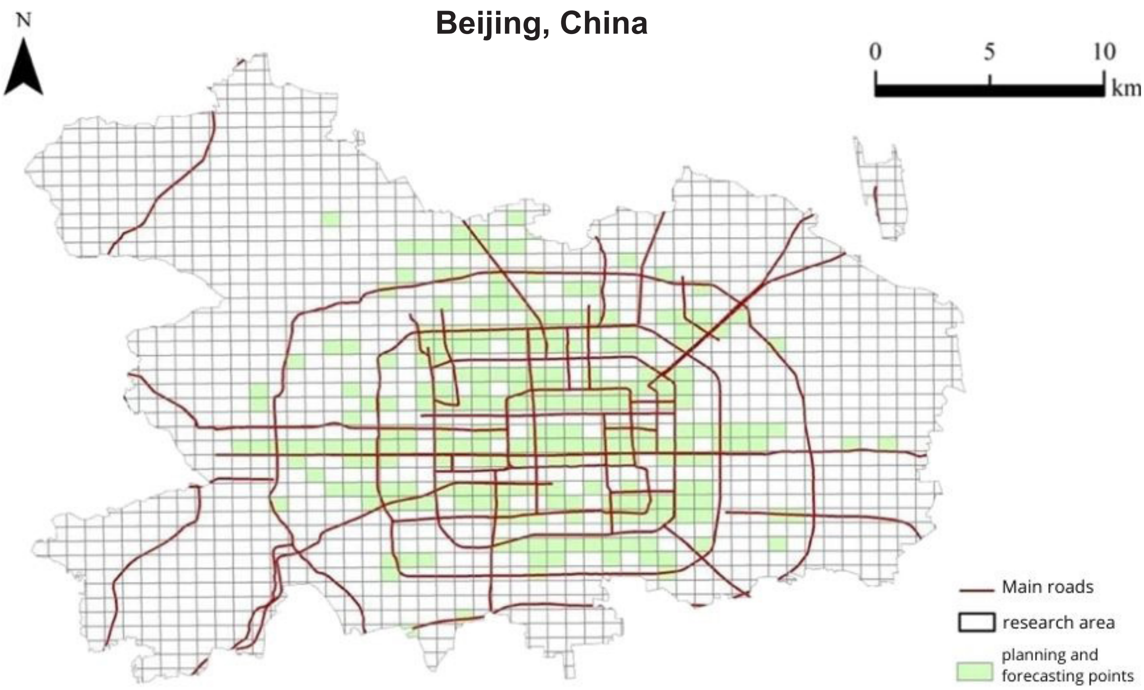

Figure 4 – Prospective sites for the location of new tourism and leisure facilities within the urban area of Beijing (green grid cells), showing main roads, study area, and planning points

Note: Map of Beijing, China

Note – compiled by the author based on the data source

Figure 3 shows, in green, the grid cells that the model identifies as promising sites for the location of new tourism and leisure facilities within the main urban districts of the city. This visualisation clearly reflects the influence of infrastructural factors in the city centre and in high-density residential areas – in particular, transport, retail, and public services – on the siting of such facilities. At the same time, the presence of green cells in some peripheral districts indicates locations that may be considered as potential areas for future tourism infrastructure development.

These results suggest that, within the framework of the WDS, tourism contributes at the macro level to reducing regional inequality, while the Beijing case illustrates, at the micro level, the spatial patterns generated by infrastructure and population density.

By combining macro-level (national and regional) WDS dynamics with micro-level (Beijing) spatial analysis, it becomes possible to discern the multi-level linkages between tourism policy and urban infrastructure. In the long term, regional policy can support tourism as a means of promoting employment and reducing poverty, whereas at the micro scale, the configuration of infrastructure, transport accessibility, and population density determines which specific districts become attractive for tourism development. This two-level perspective is crucial for data-driven planning: where national strategy designates tourism as a priority sector, cities can employ POI big data and machine-learning methods to develop concrete spatial solutions for investors and local authorities.

Conclusion

This study used a two-level analytical system. First, it assessed the impact of Western development strategies on tourism growth and regional inequality at the macro level. Second, using Beijing as a representative developed city, it demonstrated the application of POI big data and machine learning for spatial planning of tourist attractions at the micro level. Although Beijing itself is not a direct beneficiary of the WDS, the methodology developed here is directly applicable to cities in western China, offering a data-driven tool to support the strategic goals of the WDS by optimising investment in tourism infrastructure and reducing intra-regional disparities.

In this research work, the analysis of the geographical distribution of tourism and leisure facilities in the core districts of the city of Beijing has been carried out using the POI data and the CART algorithm. The findings of this research work confirm the high concentration of tourism and leisure facilities in the city center and the significant impact of infrastructural attributes – hotels, shopping centers, transportation infrastructure, and public services – as well as the distribution of the city's population. The findings of the research work can also be applied to the wider concept of smart tourism.

The model's predictive accuracy was at 83.5%, thus validating the effectiveness of ML models in analyzing tourism potential within the city space. Moreover, the model suggested the possible locations in the peripheral regions, thus being instrumental in informed decision-making regarding the location of new infrastructure development. The predictions were effective in identifying regions within the outer districts to which the boundaries of promising locations in Dongcheng, Xicheng, Chaoyang, Haidian, and Fengtai could be committed.

Therefore, the research goal has been accomplished, and the primary hypothesis has been confirmed: the hypothesis according to which the built environment of the city and the density of the city's population statistically significantly affect the geographic concentration of the points of interest of the tourism and leisure sector has been proven correct. The paper has also demonstrated the effectiveness of the combination of POI big data and the CART model in this context.

Concerning the smart tourism cities' point of view, the study has clearly shown the applications of data and machine learning algorithms regarding the spatial allocation of tourism and leisure facilities. The results of the study demonstrated that the local infrastructure and the residents' distribution are the main factors determining the location of new facilities. The paper's authors also noted the limitations of the current method's ability to provide a precise distinction of each type of facility and emphasized the necessity of a more detailed research approach in the future. In addition, it has been suggested that the next studies should combine the applications of the novel technologies of big data, the Internet of Things (IoT), and artificial intelligence.

In general, the research provides a set of findings of significance both in practice and in science concerning the development of urban tourism and the identification of efficient management strategies. The findings of the research provide novel methods of spatial pattern identification in the context of urban tourism and can be used to facilitate decision-making of city administrations and the tourism industry.

Conflict of Interest Statement

The authors declare no potential conflicts of interest regarding the research, authorship, or publication of this article.

Acknowledgments

(1) This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26198345, title: "Reducing socio-economic inequality in the regions of Kazakhstan through investment in and improvement of the organization of the healthcare system").

References

Chen, R., Yan, H., Liu, F., Du, W., & Yang, Y. (2020). Multiple global population datasets: Differences and spatial distribution characteristics. ISPRS International Journal of Geo-Information, 9(11), 637.

Croes, R., & Vanegas, M. (2008). Cointegration and causality between tourism and poverty reduction. Journal of Travel Research, 47(1), 94–103.

Demurger, S., Sachs, J. D., Woo, W. T., Bao, S., Chang, G., & Mellinger, A. (2002). Geography, economic policy, and regional development in China. Asian Economic Papers, 1(1), 146–197.

Deng, T., Hu, Y., & Ma, M. (2019). Regional policy and tourism: A quasi-natural experiment. Annals of Tourism Research, 74, 1-16.

Fan, C. C., & Sun, M. (2008). Regional inequality in China, 1978-2006. Eurasian Geography and Economics, 49(1), 1–18.

Fang, Y., Yin, J., & Wu, B. (2021). Can tourism development improve urban livability? A double machine learning approach. Journal of Destination Marketing & Management, 20, 100568.

Gao, S., Janowicz, K., & Couclelis, H. (2017). Extracting urban functional regions from points of interest and human activities on location-based social networks. Transactions in GIS, 21(3), 446-467.

Gibson, J., & Boe-Gibson, G. (2021). Nighttime lights and county-level economic activity in the United States: 2001–2019. Remote Sensing, 13(14), 2741.

Grewal, B. S. (2011). China's economic growth and its regional development strategy: The case of Western Development. Journal of Contemporary China, 20(71), 681–699.

He, D., Chen, Z., Ai, S., Zhou, J., Lu, L., & Yang, T. (2021). The spatial distribution and influencing factors of urban cultural and entertainment facilities in Beijing. Sustainability, 13(21), 12252.

Li, X., & Law, R. (2020). Network analysis of big data research in tourism. Tourism Management Perspectives, 33, 100608.

Li, X., Li, H., Pan, B., & Law, R. (2021). Machine learning in internet search query selection for tourism forecasting. Journal of Travel Research, 60(6), 1213-1231.

Li, X., Yang, T., Meng, B., Chen, S., & Zhang, S. (2025). Analysis of tourist behavior patterns and perceptions in Beijing based on user-generated content data. Applied Spatial Analysis and Policy, 18(3), 1-25.

Liu, M., & Hao, W. (2020). Spatial distribution and its influencing factors of national A-level tourist attractions in Shanxi Province. Acta Geogr. Sin, 75(04), 878-888.

Loh, W. Y. (2011). Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1), 14-23.

Lu, J. (2011). Evaluation of the Western Development Strategy: Effectiveness, challenges and the way forward. Chinese Economy, 44(5), 6–24.

Luo, Q., Liu, Y., Bi, M., Kuai, X., Tian, Q., Sun, Y., & Zhuang, S. (2024). Taste mapping: Navigating the spatiotemporal link between diet and colorectal cancer. IEEE Access, 12, 17735-17747.

National Bureau of Statistics of China. (2010). China statistical yearbook 2010. China Statistics Press.

Qian, J., Liu, Z., Du, Y., Liang, F., Yi, J., Ma, T., & Pei, T. (2021). Quantifying city-level dynamic functions across China using social media and POI data. Computers, Environment and Urban Systems, 85, 101552.

Qin, X., Wei, Y. D., Wu, Y., & Huang, X. (2023). Regional development and inequality within city regions: A study of the Yangtze River Delta, China. Geographical Review, 113(3), 359-385.

Yang, Y., Fik, T., & Li, X. (2021). Tourism and regional income inequality: Evidence from China. Annals of Tourism Research, 88, 103163.

Yi, D., Yang, J., Liu, J., Liu, Y., & Zhang, J. (2019). Quantitative identification of urban functions with fishers' exact test and POI data applied in classifying urban districts: A case study within the sixth ring road in Beijing. ISPRS International Journal of Geo-Information, 8(12), 555.

Wang, C., & Wei, Y. (2020). Does the Western Development Strategy promote enterprise investment? Evidence from China. China Economic Review, 60, 101399.

Wu, S., Wang, J., Jia, Y., Yang, J., & Li, J. (2025). Planning and layout of tourism and leisure facilities based on POI big data and machine learning. PLOS ONE, 20(3), e0298056.

Zheng, H., Huang, Y., & Wang, S. (2022). Does the Western Development Strategy promote regional economic growth? A retrospective analysis with a new econometric approach. Economic Modelling, 116, 106024.

Information about authors:

Spankulova Lazat Seitkazievna – Doctor of Economic Sciences, Professor, Al-Farabi Kazakh National University (Almaty, Kazakhstan, e-mail: spankulova.lazat@mail.ru);

Elai Erbolat – PhD Student, Abai Kazakh National Pedagogical University (Almaty, Kazakhstan, e-mail: erbolatelai@mail.ru);

Dauletkhanova Yerkezhan Ruslankyzy – PhD Student, Department of Recreational Geography and Tourism, Al-Farabi Kazakh National University (Almaty, Kazakhstan, e-mail: daulethanova_e@ mail.ru).

Авторлар туралы мәлімет:

Спанкулова Лазат Сейтказиевна – экономика ғылымдарының докторы, профессор, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан, e-mail: spankulova.lazat@mail.ru);

Елай Ерболат – PhD-докторант, Абай атындағы Қазақ ұлттық педагогикалық университеті (Алматы, Қазақстан, e-mail: erbolatelai@mail.ru);

Дәулетханова Еркежан Рүсланқызы – рекреациялық география және туризм кафедрасының PhD-докторанты, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан, e-mail: daulethanova_e@mail.ru).

Сведения об авторах:

Спанкулова Лазат Сейтказиевна – доктор экономических наук, профессор, Казахский национальный университет имени аль-Фараби (Алматы, Казахстан, e-mail: spankulova.lazat@mail.ru);

Елай Ерболат – PhD-докторант, Казахский национальный педагогический университет имени Абая (Алматы, Казахстан, e-mail: erbolatelai@mail.ru);

Дәулетханова Еркежан Рүсланқызы – PhD-докторант кафедры рекреационной географии и туризма, Казахский национальный университет имени аль-Фараби (Алматы, Казахстан, e-mail: daulethanova_e@mail.ru).

Received: October 1, 2025
Accepted: December 02, 2025