IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 July 2023, accepted 9 August 2023, date of publication 17 August 2023, date of current version 24 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3305957

RESEARCH ARTICLE

Modeling the Time Spent at Points of Interest Based on Google Popular Times

ALI JAMAL MAHDI [1,2], TAMÁS TETTAMANTI [3], AND DOMOKOS ESZTERGÁR-KISS [1]

[1]Department of Transport Technology and Economics, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, 1111 Budapest, Hungary
[2]Department of Construction and Projects, University Headquarter, University of Anbar, Ramadi 31001, Iraq
[3]Department of Control for Transportation and Vehicle Systems, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, 1111 Budapest, Hungary

Corresponding author: Ali Jamal Mahdi (alijamalmahdi@edu.bme.hu)

This work was supported in part by the National Research, Development and Innovation (NRDI) Office, Hungary, and the ERA-NET COFUND/EJP COFUND Program with co-funding from the European Union Horizon 2020 Research and Innovation Program under Project 2019-2.1.7-ERA-NET-2021-00019; and in part by the János Bolyai Research Fellowship of the Hungarian Academy of Sciences under Grant BO/00090/21/6.

ABSTRACT Location-based applications are increasingly popular as smartphones with navigation capabilities are becoming more prevalent. Analyzing the time spent by visitors at Points of Interests (POIs) is crucial in various fields, such as urban planning, tourism, marketing, and transportation, as it provides insights into human behavior and decision-making. However, collecting a large sample of behavioral data by using traditional survey methods is expensive and complicated. To address this challenge, this study explores the use of crowdsourcing tools, specifically Google Popular Times (GPT), as an alternative source of information to predict the time spent at POIs. The research applies a robust regression model to analyze the data obtained from GPT. The popularity trends of the different POI categories are used to indicate the peak hours of the time spent in the city of Budapest. Non-spatial parameters such as the rating, the number of reviewers, and the category of the POIs are utilized. Furthermore, a Geographic Information System (GIS) is applied to extract the spatial parameters such as the security and safety levels, the availability of car parking, and public transport (PT) stations. The robust linear models are statistically significant based on the p-values, thus indicating a strong relationship between the independent variables and the time spent at POIs. The weekday and weekend models present 69.5% and 73.9% of the variance in the time spent at POIs, respectively. Furthermore, it is demonstrated that the visitors' behavior is strongly affected by the category of the POIs variable. This study shows how GPT can be utilized to better understand, analyze, and forecast people's behavior. The solution presented in this study can serve as an essential support of activity-based models, where the time spent is a crucial parameter for scheduling and optimizing activity chains.

INDEX TERMS Point of Interest, Google popular times, GIS, time spent, regression model.

I. INTRODUCTION

Travelers' behavior and time patterns have typical forms depending on several factors, where an individual's activity pattern is generated from the decisions related to location choices and the time spent on activities. Considering the transportation-related context, the time spent is the time allocated by an individual to a specific activity at a specific Point of Interest (POI), such as a restaurant, a museum, or a store [1]. In fact, the time spent on an activity is mostly related to non-work-related activities. People allocate time according to constraints, preferences, features related to the POIs, and travel circumstances [2].

Collecting information about locations by traditional methods requires a huge amount of effort and cost, while location-based services (LBSs) provide a suitable way to gather the relevant information [3]. Mobile phones have emerged as a source of information because they provide a

The associate editor coordinating the review of this manuscript and approving it for publication was Giacomo Fiumara.

vast amount of spatiotemporal data [4]. Due to the widespread usage of smartphones, it is possible to identify various locations of activities, which meet individuals' interests, even in unfamiliar areas [5]. The spatial and temporal patterns of an aspect can be identified by extracting information with the help of sensing devices, where individuals share data [6]. Another study confirms that identifying people's patterns and trends is difficult to detect by using traditional research methods [7]. According to Junglas and Watson [8], Google Popular Times (GPT) is both an LBS and a data source, which provides insights into the popularity of the places. As a result, data on the number of visitors and the length of their visits can be linked with user reviews and other information from third-party sources [9]. However, GPT cannot indicate the actual number of visitors, and the sampling bias is an issue because the collection of the data is based on using smartphones [10]. The aggregated databases should alleviate these shortcomings (e.g., POIs for a whole city), where GPT is an appropriate source for creating these databases.

In the context of transportation planning, modeling the time spent is crucial. These models may be used to assist planning and decision-making related to transportation systems. In addition, modeling the time spent is relevant for personal scheduling. An individual's activity chain consists of two main sections: the travel time between the activities and the time spent on an activity. Travel time has been handled by several studies [11], [12], [13], and [14]. However, the time spent by individuals at POIs has not been extensively studied in previous papers.

It is worth mentioning that the current study utilizes statistical methods, particularly robust regression models, to explore the relationship between the time spent at POIs and the relevant variables. The reasons behind using these models instead of other methods (such as machine learning techniques) are the followings; First, robust regression models can be applied to various problems as they contribute to understanding the global impacts of several features [15]. Second, the efficiency of machine learning algorithms is highly influenced by the data including outliers [16]. Therefore, the current study avoids applying machine learning algorithms, which are highly sensitive to the data containing outliers, and the presence of outliers in the collected data of the current study could significantly influence the findings.

So far, no study has developed a comprehensive model for accurately predicting the time spent by individuals at POIs by using GPT data. Therefore, this study develops comprehensive models that incorporate GPT data and considers realistic factors. Furthermore, previous studies have not adequately addressed the temporal variations in the time spent at POIs, specifically on weekdays and weekends. Accordingly, this study aims to fill this gap by proposing robust models for predicting the duration spent at POIs on weekends and weekdays while taking the following realistic factors into account: rating, number of reviewers, category of POIs, security and safety levels, as well as the availability of car parking and public transport (PT) stations. By addressing these research gaps, this study seeks to enhance the understanding of individuals' time allocation patterns and to provide valuable insights into transportation planning and personal scheduling.

The rest of this paper is structured as follows. The literature review is presented in Section II. Section III explains the methods applied in this research. Section IV presents the results, while Section V discusses policy implications, possible limitations, and planned future works. Finally, the conclusion is presented in Section VI.

II. LITERATURE REVIEW

Since individuals have limited time during the day, the travel time and the time spent on each activity should be examined precisely. Several previous studies have analyzed the time of the journey, where traffic circumstances and sociodemographic factors are utilized as influential factors. Many studies handle the travel time of the PT mode [17], [18], while other research works focus on predicting the travel time of private vehicles [11], [19]. However, there are merely few studies that develop models for estimating the time spent on activities.

There are some attempts to analyze visitors' behavior and to model the time spent. A study uses the Poisson regression model to estimate the frequency and time allocation of leisure activities associated with individual, household, and spatial factors [20]. The results indicate that the spatial factors have less impact compared with other adopted factors. However, the spatial factors with the lack of transportation modes might discourage individuals from participating in their desired activities [21]. Another study [22] analyzes the factors that influence the time spent on shopping activities on workdays, where spatial environment variables and time restrictions are studied through Hazard-based methods. The findings show that time restriction has a significant impact on the amount of the time spent on shopping. These findings emphasize the significance of spatial factors and time constraints in understanding the time allocation at specific activities. The cultural tourists tend to spend more time at heritage destinations, while another study mentions that several factors can influence the length of the stay including destination attributes, travelers' characteristics, and motivations [23]. Visitor demographics, such as age, gender, and income, are found to affect the time spent, especially younger visitors and those with higher income tend to spend more time at destinations [24]. The quality of the visitor experience, including factors such as customer service, cleanliness, and safety, has been identified as important in influencing the time spent at destinations [25]. In addition, a study finds that the duration of museum visits is influenced by several factors, including visitor characteristics, the exhibition content, and the museum environment [26]. While the previous studies focus on specific types of attractions, they provide insights into the factors that can influence the time allocation at different POIs. The effect of changing the number of workdays on individuals' lifestyle has been examined by [27]. The outcomes of the SEM models reveal a strong relationship

TABLE 1. Summary of the previous attempts to model the time spent at POIs.

Reference	Activity type	Factor	Data collection	Study area
[33]	Various POIs	Spatial factors, such as lack of transportation modes	Traditional	Sydney, Australia
[22]	Shopping	Time restrictions	Traditional	Bristol and Eindhoven
[23]	Heritage destinations	Immersive experiences	Innovative techniques	Hong Kong
[24]	Various POIs	Visitor demographics, such as age, gender, and income	Traditional	Singapore
[25]	Various POIs	Quality of the visitor experience, such as customer service, cleanliness, and safety	Traditional	Hawaii, USA
[26]	Museums	Visitor characteristics, exhibition content, museum environment	Innovative techniques	United States
[27]	Leisure activities	Sociodemographic characteristics, time spent on leisure activities, travel behavior	Traditional	Seoul, Korea
Current study	Various POIs	Rating, number of reviewers, POI type, PT station availability, car parking availability, security level, and safety level	Innovative techniques	Budapest, Hungary

among the sociodemographic characteristics, the time spent on leisure activities, and the travel behavior. A study aims to investigate the relationship between the destination image and tourist behavior. The impact of the city destination image, distance, family income, perceived expensiveness, and age has been examined. The findings highlight the influential role of the destination image, along with the other identified variables, in shaping travel patterns [28]. Another study presents a Personal Navigation System (PNS) designed to efficiently guide tourists through multiple destinations. The PNS allows tourists to specify their desired arrival and stay times, as well as their preference degree for each destination. The system calculates a route that satisfies tourists' requirements and navigates them accordingly [29]. Analyzing visitors' time-based activities at destinations has been handled by [30], who examines the visitor behavior by using a time block analysis approach. Data are collected through face-to-face interviews by using a diary-type questionnaire in 13 ski centers in Greece. The study classifies time periods and describes the visitor flow and behavior within different time blocks throughout the day. Expenditure patterns are identified in relation to specific time blocks indicating preferences for certain products and services. Generally, merely one activity type is examined in previous studies, where traditional survey methods with some limitations are used to collect the relevant data. Therefore, this study provides added value by using a robust method and collecting data from larger geographical areas via GPT, as well as modeling the time spent on several activity types. Current study aims to leverage the vast amount of big data available on Google as an additional data source for collecting information and modeling the time spent at POIs.

The large collection of geospatial datasets available through the Internet is crucial for urban modeling as it enables the analysis of individuals' spatial and temporal patterns [31]. A recent paper showing the applicability of GPT in predicting travel behavior is conducted by [10], where the restaurants in the city center of Munich are used to examine the actual consumer visit behavior by considering the customer reviews, the timing effects, the quantity of the uploaded images, and the pricing information. The researchers conducted bivariate linear regression and correlation analyses finding that the number of uploaded photos and ratings correlate and can be used to estimate the number of tourists. Additionally, a negative effect of the visitor numbers on the visit duration is demonstrated. A recent study on the popularity of POIs by [32] examines the activity and demand trends, where the scholars demonstrate that POI check-ins can be a valuable source of data during disruptive events. The findings demonstrate the impact of variables, such as parking space and PT stop distance, on the popularity of POIs. Current study exploits the popularity of POIs from another viewpoint. The focus is on estimating the time spent at POIs in Budapest, where the time spent at POIs is considered as the continuous dependent variable, while robust regression models are applied. Table 1 presents a summary of the previous attempts to model the time spent at POIs.

Previous studies establish a foundation for current research by discussing the time spent on activities, visitor behavior, and factors influencing time allocation at POIs. It highlights the importance of considering various factors, such as spatial variables and destination characteristics, in understanding and modeling the time spent at POIs.

III. METHODOLOGY

In this paper, the GPT is examined as a reliable source to model the time spent at POIs. This study focuses on predicting relationships between the time spent and the considered factors (i.e., spatial and non-spatial factors). The methodology is based on two fundamental steps: the data collection and

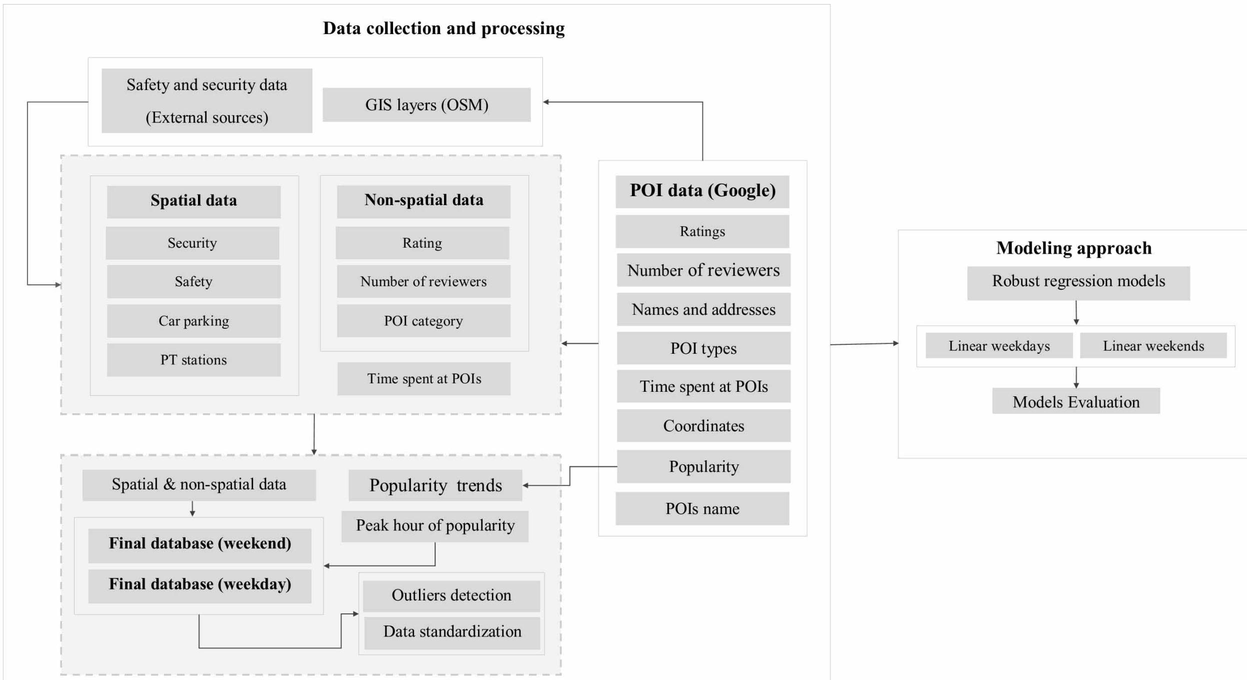

FIGURE 1. **The steps of the method to discover the time spent at POIs.**

the modeling approach. In the first step, GPT is used to collect data at various POIs on weekdays and weekends in Budapest city. The second step involves developing a robust regression model to estimate the time spent at POIs on weekdays and weekends. The proposed steps of the method are illustrated in Figure 1.

A. DATA COLLECTION

This study deals with two types of data: the non-spatial data of POIs and the spatial data around POIs. The non-spatial data are obtained via Python code [32] for a period of one month in July 2021. The spatial data are acquired by using Open Street Map (OSM), which is based on QGIS software [34]. A sample of 2291 POIs grouped into 73 types of POIs is obtained. The collected data include variables such as the name and type of POIs, the ratings, the number of reviewers, the coordinates, the time spent by individuals at POIs, and the popularity. The spatial variables include safety, security, car parking availability, and the PT stations around POIs. This research applies a simple approach to group the POIs into categories by using their names. According to [3], the names of POIs expose the category of the POIs. In addition, the spatial-temporal POI attractiveness is utilized to group the POIs, where those POIs that have the same time windows are grouped into the same category [35]. This concept is applied to split the POIs between the tourist attraction and the entertainment categories. It is worth noting that this study uses dummy variables to measure these predictors, where the following five categories are used to group the POIs:

- ***Public facilities:*** *including public or private offices, hospitals, agencies.*
- ***Dining:*** *referring to restaurants and all locations related to eating.*
- ***Shopping:*** *referring to malls, retail stores.*
- ***Entertainment:*** *including indoor leisure activities with late closing time, e.g., clubs.*
- ***Tourist attractions:*** *including historical places, museums, galleries, and outdoor leisure activities.*

The adopted non-spatial parameters include the rating and the number of reviewers. The rating represents the individuals' reviews of each POI on a scale (i.e., 1 star: very bad and 5 stars: very good). A higher rating indicates a more favorable attitude toward the relevant component [36]. The number of reviewers on Google represents the users who leave reviews with positive or negative reviews based on their experience.

Regarding the spatial parameters, a GIS-based walking distance analysis is used to measure the potential accessibility, which is determined by the distance between the car parking spots or PT stations and the destinations. The availability of car parking or PT stations is categorized into three variables: high, medium, and low indicated with the numeric values 3, 2, and 1, respectively. Buffer analysis by using GIS software is used to extract the values based on the car parking and PT stations availability around the POIs. A buffer of 400 m or less is used for POIs with high category, a buffer between 400 m and 600 m is used for the medium category, while the POIs 600 m away from the car parking or the PT stops are indicated by low category. Typically, transportation agencies utilize a 400 m walking distance when assessing

TABLE 2. Description of the adopted spatial and non-spatial parameters.

Parameter	Spatial parameter	Type	Scale
Rating	No	Continuous	-
Number of reviewers	No	Continuous	-
Category	No	Nominal	Dummy variable (1 for the used category type and 0 for others)
PT stations availability	Yes	Ordinal	3 = High level 2 = Medium level 1 = Low level
Car parking availability	Yes	Ordinal	3 = High level 2 = Medium level 1 = Low level
Security	Yes	Ordinal	From 1 (good condition) to 10 (bad condition)
Safety	Yes	Ordinal	From 1 (good condition) to 10 (bad condition)

accessibility. Reference [37] mention in their study that the typical walking distance between dwellings and the local PT is less than 600 meters. For these reasons, current study uses the buffers of less than 400 m and more than 600 m as an indicator for favorable and unfavorable conditions. It is worth mentioning that merely the off-street parking spaces are considered in current study. Similarly, the PT stations are represented exclusively by the metro and tram stations. The reason behind this consideration is that the metro and tram systems are more significant and reliable in ensuring PT accessibility [38]. For the safety and security variables, the attribute data on district level are obtained from [39]. The levels of security and safety are represented by the number of crimes per 1000 inhabitants and the number of accidents per 100.000 inhabitants for each district. A scale from 1 to 10 is used to rank the level of the safety and security variables. The high level is represented by the value 1 (i.e., the lowest number of crimes or accidents), while 10 (i.e., the highest number of crimes or accidents) refers to a low level of safety or security. The adopted non-spatial and spatial parameters are detailed in Table 2.

B. MODELING APPROACH

The popularity of POIs highlighted in this study varies as the time of the day changes. The average number of visitors for the five POI categories is used to model the demand trends at the POIs. Therefore, the time spent is estimated according to the popularity hours of the POIs. The collected data contain the time spent for each hour, but the calculations are realized for the maximum popularity (i.e., busyness). As an illustrative example, the popular times at the Fisherman's Bastion, which is a tourist attraction POI in Budapest, are demonstrated in Figure 2. It is shown that the visitors' maximum popularity for a given day is at 12:00. This period is utilized to model the time spent for the tourist attraction category.

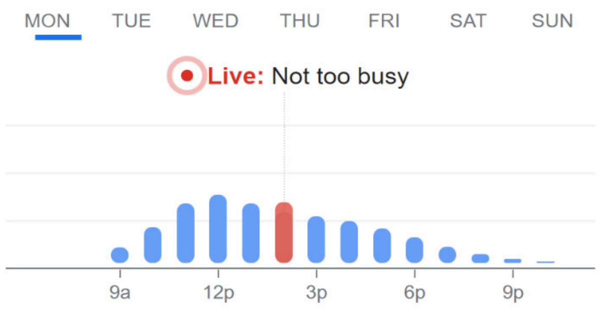

FIGURE 2. Popular times at one of the tourist attractions in Budapest [40].

As POIs are grouped into categories, the patterns of popularity are calculated for each category. Based on this, the most crowded hour is indicated to study the relationship between the time spent and the considered parameters on weekdays and weekends. After identifying the relevant time spent period for each category based on the peak crowded hour, robust linear regression models are utilized. It is worth mentioning that the POIs of public facilities type are removed from the weekend database because these POIs are closed on weekends.

A robust multiple linear regression model is applied to investigate the relationship between the time spent and the considered independent variables. The purpose of the linear regression analysis is to figure out how a dependent variable is related to a group of regressors in a linear way [41]. The linear regression model is denoted by Equation (1).

$$Y_i = \alpha + \beta_1 X_i + \ldots\ldots\ldots \beta_p X_p \quad (1)$$

where Y_i is an $n\times1$ vector containing the values of the dependent variable, and X_i is an $n\times p$ matrix having the values for the p regressor values. The term β contains unknown regression parameters that should be determined, and α is the intercept. The linear regression equation fits the predicting model to an observed dataset for forecasting purposes. With the new additional observed values of X, the fitted model can be used to forecast the value of Y.

The outcomes are obtained by conducting the regression analysis concerning whether the independent variables in the robust model predict the time spent significantly during the weekdays and weekends. The dependent variable is the time spent for the POI categories, which should be mapped to a number of variables. This can be expressed mathematically for the i^{th} POI category on weekdays or weekends, as shown in Equation (2).

$$\mathrm{T(weekday,\ weekend)_i} = \mathrm{f\,(CA_i, RN_i, RA_i, SA_i, SE_i, CP_i, PT_i)} \quad (2)$$

where:

T_i is the time spent in minutes of the i POI on weekdays or weekends,

CA_i is the POI category, dummy variable {dining (DI), entertainment (EN), public facilities (PF), shopping (SH), and tourist attraction (TA)},
RN_i is the number of reviewers of $\mathrm{POI_i}$,
RA_i is the rating of $\mathrm{POI_i}$,
SA_i is the safety level of $\mathrm{POI_i}$,
SE_i is the security level of $\mathrm{POI_i}$,
CP_i is the availability of car parking around $\mathrm{POI_i}$,
PT_i is the availability of PT station around $\mathrm{POI_i}$,

For the linear regression, it is feasible to fit the dependent variable by using the estimated parameter $\hat{Y}$ and calculate the residuals $\hat{R}$, as demonstrated in Equation (3).

$$\left.\begin{array}{c}\hat{\mathrm{Y}}=\mathrm{X}\hat{\beta}\\ \hat{\mathrm{R}}=\mathrm{Y_i}-\hat{\mathrm{Y}}_{\mathrm{i}}\\ \text{for } 1\leq \mathrm{i}\leq \mathrm{n}.\end{array}\right\} \tag{3}$$

The ordinary least-squares (OLS) method is one of the most widely applied approach for estimating $\hat{\beta}$. This method is used to determine the parameters by minimizing the sum of the squared residuals (i.e., the difference between the actual and fitted values of the dependent variable), which are very sensitive to the outliers, especially those that arise in high leverage situations [42]. It is worth mentioning that vertical outliers arise in the database of current research. According to [42], vertical outliers are observations with extremely large values in the response variable, whereas leverage points are data with extremely large values in the explanatory variable. Because outliers have an impact on the classical regression, approaches that are insensitive to them are required. Hence, full robustness can be attained by two typically used methods: diagnostic approach and robust procedures [43].

Diagnostic techniques attempt to identify odd observations by using diagnostic statistics and remove them from the data. Afterward, traditional procedures, such as the OLS, are performed to clean the dataset. This method works well for simple data, or when there are solely one or two outliers, but it is inefficient for large numbers of outliers in a complex dataset [44]. Since the outliers are massively present in the database of current study, it is not efficient to use a diagnostic approach where it affects the data, especially for the tourist attraction category. Therefore, the robust regression approach is used instead of the OLS to model the time spent at POIs on weekdays and weekends. Although $\hat{\beta}$ can be calculated in different ways, the purpose is always to reach close to the observed value by diminishing the magnitude of the residuals. To achieve this, a robust model is applied for estimating $\hat{\beta}$, as shown in Equation (4) [45].

$$\hat{\beta}=\mathrm{argmin}_{\beta}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\beta)^2\mathrm{R_i} \tag{4}$$

where $\mathrm{R_i}(\beta) = \mathrm{Y_i} - \beta_\circ - \beta_1\mathrm{X_{i1}} - \ldots\ldots\ldots\beta_\mathrm{p}\mathrm{X_{ip}}$ for $1 \leq \mathrm{i} \leq \mathrm{n}$.

The satisfactory R^2 (i.e., goodness of fit) value alone is not adequate to accept the proposed model, it has to be validated, as well. Several methods can be used to validate regression models. Data splitting or cross-validation is where a part of the data is used to compute the model coefficients, and the remaining data are utilized to determine the prediction accuracy of the model [46]. Usually, more observations should be used for estimation than validation. Therefore, about 15% of the final data are used for validating the developed models [47]. The relative error (RE) is a fundamental indicator for measuring the validation of the models, which can be calculated by Equation (5) [48]. For transparency and to provide more comprehensive results, another error calculation, i.e., the root mean square error (RMSE), is considered and applied by using Equation (6) [49].

$$\mathrm{RE\%}=\frac{\mathrm{PT_{actual}}-\mathrm{PT_{prediction}}}{\mathrm{PT_{actual}}} \tag{5}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{PT_{prediction}}-\mathrm{PT_{actual}})^2} \tag{6}$$

where PT_{actual} is the time spent of the cross-validation sample. $PT_{prediction}$ is the time spent estimated by the model, and n is the number of errors. In both RE and RMSE, the directions of the errors are ignored. However, RMSE always provides a positive number and a real solution because the expression under the root symbol is non-negative.

IV. RESULTS

A. THE CASE STUDY AND SPATIAL DISTRIBUTION

Budapest is the capital of Hungary integrating political, economic, administrative, cultural, and symbolic functions accounting for 18% of the total population of the country. Moreover, Budapest is the most popular Hungarian tourist destination visited intensively during the whole year [50]. The city has a total area of 525 km^2 and a population of ca. 1.7 million inhabitants. The Hungarian capital is divided into 23 districts, and each district has its own set of economic, social, and cultural features [51]. A variety of POIs including workplaces, residential areas, and leisure time venues, such as restaurants, cinemas, and tourism attractions can be found in the city [52].

As mentioned in the methodology section, POIs are grouped into five categories: public facilities, dining, shopping, entertainment, and tourist attraction. Figure 3 shows the spatial distribution of these categories in the study area. 40% of the POIs belong to the shopping category, and with 22%, dining is the second category. The percentage of the tourist attraction (14%), entertainment (13%), and public facilities categories (11%) is approximately the same.

Figure 4 shows the results of the spatial analysis. It is clear from the results of the spatial analysis that car parking places for the majority of the POIs are available. While the spatial analysis results of the PT station reveal that 70% of the POIs reachable by PT stations with a walking distance of less than 600 meters. The two bottom maps illustrate the results of the safety and security values. As it can be observed, the highest level of safety is found in Districts 23 and 15. Thus,

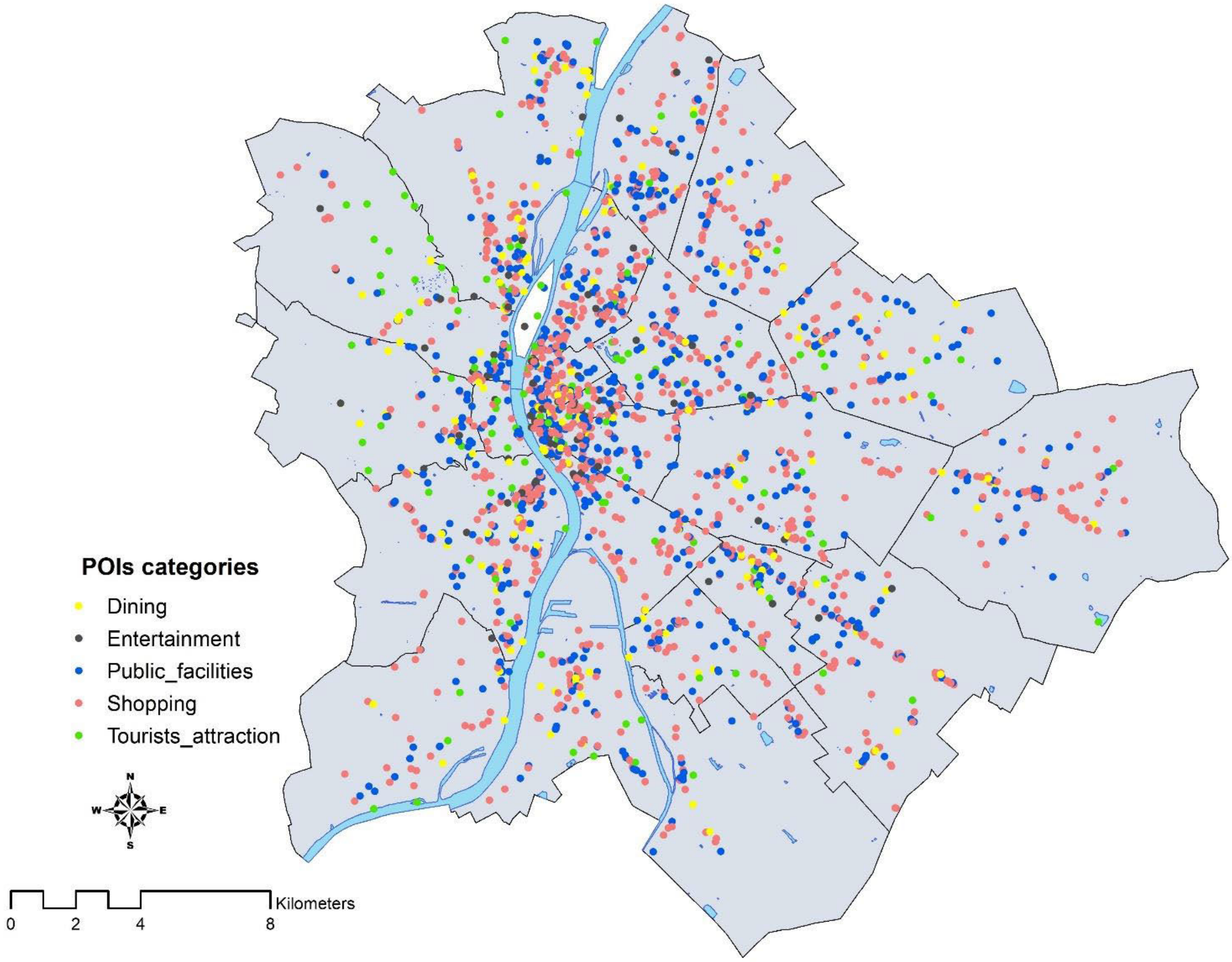

FIGURE 3. The spatial distribution of the POIs categories.

the POIs located within these districts are represented by the best safety condition (i.e., 1 in this case represents the lowest number of accidents). On the other hand, the POIs located in District 17 retain the lowest condition, and these POIs are indicated by the numeric value 10 (i.e., the highest number of accidents). In the same vein, for the security variables, District 5 has the lowest level of security; as a result, all POIs in this district have the numeric value 10, which represents the highest number of crimes. However, the POIs found in Districts 17 and 16 are represented by adequate conditions regarding the security level.

B. POPULARITY TRENDS

The average historical popularity of the five POI categories on a weekly and hourly basis is shown in Figures 5 and 6, respectively. One week is a typical time frame for everyday travel behavior because it encompasses both weekdays and weekends, and many routines are repeated weekly [53]. Tourist attraction POIs have a noticeable peak between 12:00 and 13:00, where the trends are similar for weekdays and weekends. However, it is obvious that the number of tourists on the weekend is larger than on weekdays. The peak period is chosen to study the time spent at POIs of tourist attraction destinations. For POIs within the entertainment category, it is clear that they have a significant peak during the evening hours particularly on weekends when the number of visitors is approximately four times compared with weekdays. Moreover, weekends for the tourist attraction category present a higher demand than weekdays. The dining and the shopping categories have similar trends between 10:00 and 20:00, but the number of visitors for the shopping category is larger than for dining. On weekends, the dining category exhibits two distinct peaks, which can be observed due to the increase in the number of visitors during lunch and late-evening hours. The pattern is relatively absent for the public facilities category as most of these institutions in Budapest are closed on weekends. The slight pattern appears on weekends due to the few visitors of hospitals within the public facilities

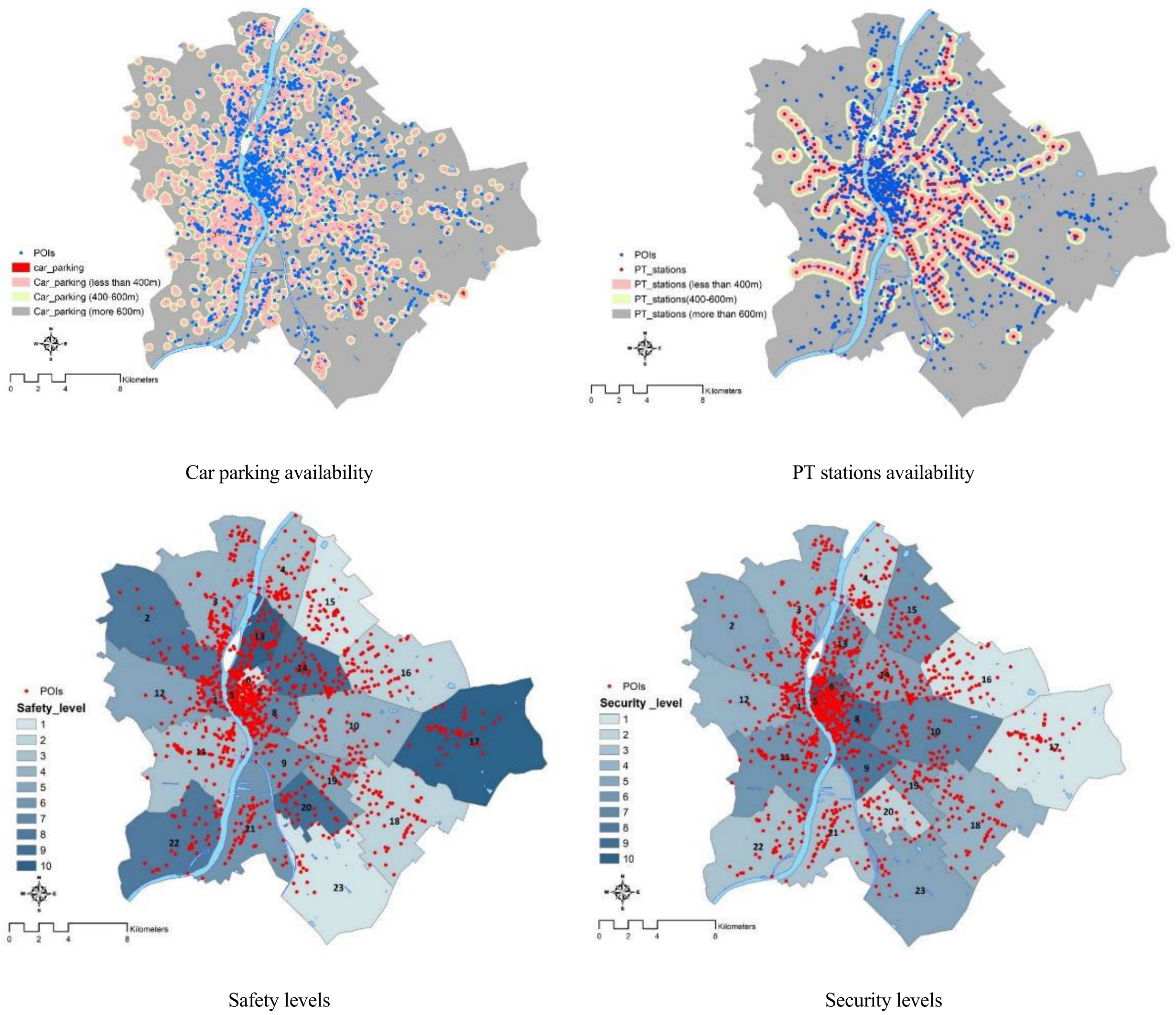

FIGURE 4. The results of the spatial analysis.

category. There is a dramatic change in the number of visitors on weekdays and weekends, where the figure represents a rise in the GPT in case of weekdays more than four times compared with weekends. Other POIs show similar patterns on weekdays and weekends with slightly higher GPT on weekdays. Moreover, two peaks of the GPT can be found. It can be seen that the number of visitors is higher on weekends for the tourist attraction and entertainment categories unlike the POIs within the public facility category. However, the shopping and the dining categories have approximately the same trends, where there is no change in the number of visitors on weekdays and weekends. After identifying the peak hours for each category on weekdays and weekends, it is straightforward to pick up the relevant time spent from the database for each day. Accordingly, the models can be applied to the prepared database to model the time spent. It is important to highlight that Equation (4) is applied for the linear regression representing the use of the robust approach instead of the OLS.

C. MODELING CONSIDERATIONS

Violin plots are used to compare the distribution of the time spent on weekday and weekend across the POIs. The plot consists of two main parts: the box plot and the density plot [54].

The box plot within the violin plot provides a summary of the data distribution. It includes five main values:

1. Minimum: The smallest observed value in the data.
2. First quartile (Q1): The value below which 25% of the data fall. It represents the lower end of the ''box'' in the plot.
3. Median: The middle value of the data. It represents the horizontal line within the ''box.''
4. Third quartile (Q3): The value below which 75% of the data fall. It represents the upper end of the ''box'' in the plot.
5. Maximum: The largest observed value in the data.

In this case, the box plot details for weekdays and weekends are presented in Table 3. For weekdays, the minimum value

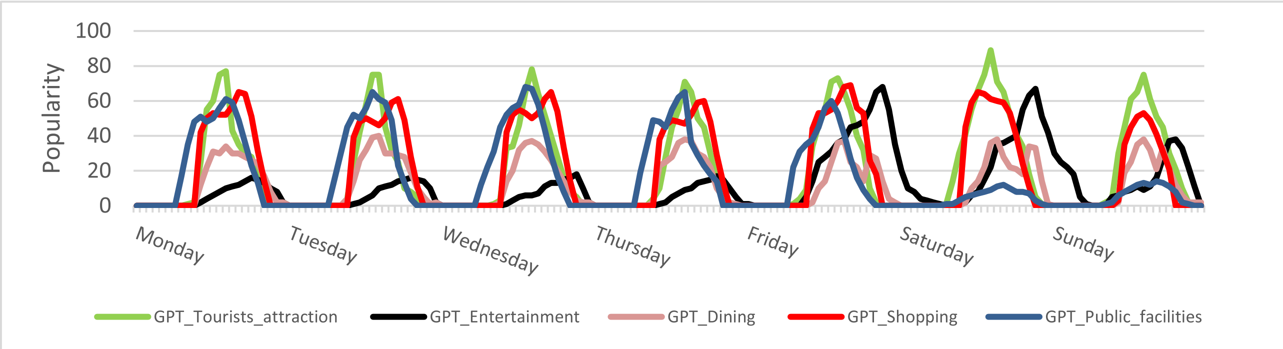

FIGURE 5. **Weekly popularity trends of the POI categories.**

TABLE 3. **The number of boxplots for weekdays and weekends.**

Boxplot details	Weekdays	Weekends
Minimum	22	18
First quartile (Q1)	23	27
Median	23	27
Third quartile (Q3)	33	37
Maximum	149	188

of the time spent is 22, the first quartile is 23, the median is 23, the third quartile is 33, and the maximum is 149. On weekends, the corresponding values are 18, 27, 27, 37, and 188, respectively.

The density plot, which forms the "violin" shape, shows the distribution of the data on weekdays and weekends (Figure 7). It represents the probability density of the different values of the time spent. A wider section of the plot indicates a higher density of the data points in that range. Based on the violin plot, it is evident that there are vertical outliers present in the collected data. Vertical outliers are data points that lie far away from the main body or the central distribution of the data. The presence of the vertical outliers can be observed by noting the maximum values of the time spent on weekdays (149) and weekends (188), which are considerably larger than the third quartile values (Q3) of 33 and 37, respectively. These extreme values indicate the presence of the outliers in the dataset. To handle the presence of the vertical outliers, a robust regression model can be used instead of the OLS method. Robust regression techniques are designed to minimize the influence of the outliers on the model estimation and provide more reliable results, particularly in the presence of extreme values.

D. ROBUST MODEL RESULTS

In Table 4, the first column represents the independent variables, such as the category of POIs (i.e., a dummy variable), the number of reviewers, the rating of POIs, safety, security, car parking, PT station availability around POIs, and the intercept. The second column is the abbreviation of the variables, and the value of the variable coefficient (Coef.) is given in the third and seventh columns, where these values reflect the significance of the independent variables. The variable with the highest coefficient value is considered the most important independent factor regardless of the sign. The fourth and eighth columns are the standard error (St. Err.). In the fifth and ninth columns, the t-value evaluates how much the discrepancy differs from the variation in the sample data. In the sixth and last columns, the p-value is the significant value or the probability, which expresses the significance of the coefficient at a confidence level of 95%. The test hypothesis is rejected if the test result is statistically significant (i.e., p-value less than or equal to 0.05). However, the variable is insignificant if the p-value is higher than 0.05.

The table indicates that the robust linear model is significant for both weekdays and weekends based on the p-values. In other words, there is a strong relationship between the independent variables and the dependent variable (i.e., time spent) in both models. The degree to which the model predicts the variance in the dependent variables (R^2) is found to be $R^2 = 0.695$, which means that 69.5% of the variation in the dependent variable can be explained by the independent variables in the weekday model. This suggests that the model has a reasonably good fit as it captures a significant portion of the variance in the data. For the weekend model, $R^2 = 0.739$, which means that 73.9% of the variation in the dependent variable can be explained by the independent variables in the weekend model. Overall, the lower RMSE and %RE values in both models indicate that the predictive model is more accurate for the weekend data than for the weekday data.

Regarding the coefficients, based on the statistical outcomes all the independent variables are significantly predictive of the time spent on weekday and weekend according to the p-value (i.e., p-value less than 0.05), except for the number of reviewers variable in the weekday model. Thus, it can be determined from these coefficients that the models predict the dependent variable accurately. While examining the contributions made by the independent variables in the linear model, it is observed that the adopted factors have approximately similar impacts on the time spent for both models. The category type (Ci) for both models has a larger impact on the dependent variable than other variables, where the tourist attraction category is the most influential type with the value of 17.53 and 16.57 for the weekdays and weekends

Popularity trends of the public facilities category

Popularity trends of the dining category

Popularity trends of the shopping category

Popularity trends of the entertainment category

Popularity trends of the touristic attraction category

FIGURE 6. **The average popular time trends of the POI categories on weekends and weekdays.**

models, respectively. The p-value for the number of reviewers in case of the weekdays variable is greater than 0.05. This means that there is no statistically significant relationship between the number of reviewers and the time spent at POIs on weekdays. On the other hand, if the p-value for the number of reviewers on weekends is significant but with a zero value, it means that the number of reviewers is not an adequate predictor of the time spent at POIs on weekends. The remaining categories have approximately similar influences. The availability of car parking on a weekdays provides values slightly larger than weekends. This indicates that people are more satisfied with car parking facilities on weekdays than on weekends. Furthermore, it is found that the rating has an impact on the time spent, where increasing the rating will increase the time spent at POIs. Without a doubt, safety and security factors negatively impact the time spent, and this effect is greater on weekends than on weekdays. This implies that people may be more concerned about safety and security issues when visiting POIs on weekends, which may cause them to spend less time at these locations. This could be due to such factors as increased crowds, higher crime rates, or a lack of security measures on weekends. The intercept values in the weekday and weekend models are 8.46 and 11.47, respectively.

V. DISCUSSION

The findings are relevant for transport planners and researchers working on scheduling activities, where the time

TABLE 4. The robust regression results on weekdays and weekends.

		Robust regression model (Weekday)				Robust regression model (Weekend)			
Variable	Abbr.	Coef.	St.Err.	t-value	p-value	Coef.	St.Err.	t-value	p-value
Category type:	C_i	-	-	-	-	-	-	-	-
Dining	DI_i	8.95	0.35	25.23	0.00	11.13	4.28	2.60	0.01
Entertainment	EN_i	9.00	0.48	18.61	0.00	11.14	4.29	2.60	0.01
Public facilities	PF_i	7.71	0.24	31.72	0.00	-	-	-	-
Shopping	SH_i	7.64	0.23	33.43	0.00	10.66	4.27	2.50	0.01
Tourist attractions	TA_i	17.53	0.45	38.68	0.00	16.57	4.28	3.87	0.00
Number of reviewers	RN_i	0.00	0.00	0.42	0.68	0.00	0.00	2.72	0.01
Rating	R_i	2.08	0.20	10.39	0.00	1.35	0.21	6.55	0.00
Safety	SA_i	-0.17	0.04	-3.98	0.00	-0.30	0.06	-5.01	0.00
Security	SE_i	-0.39	0.06	-6.66	0.00	-0.32	0.05	-6.34	0.00
Car parking	CP_i	2.90	0.19	15.47	0.00	2.52	0.18	14.00	0.00
PT stations	PT_i	2.06	0.18	11.59	0.00	2.86	0.18	16.29	0.00
Constant		8.46	0.57	14.83	0.00	11.47	4.39	2.61	0.01
Number of observations		1992				1324			
R-squared		0.695				0.739			
P-value		0.00				0.00			
%RE		9.07				5.52			
RMSE		4.11				3.37			

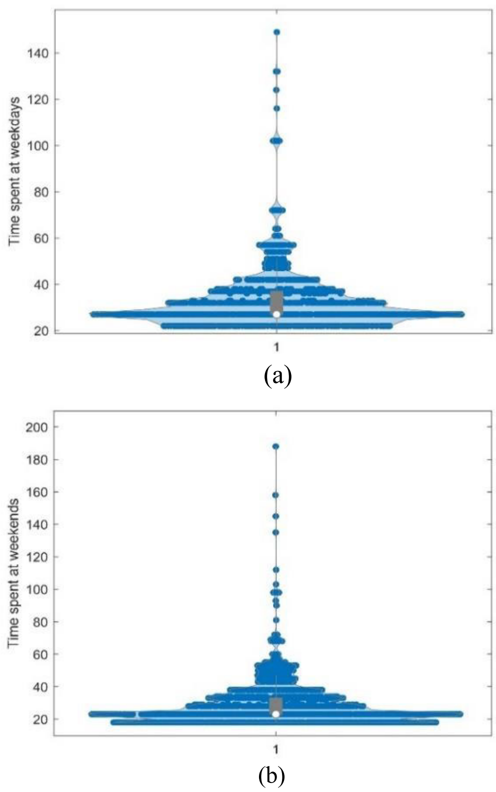

FIGURE 7. Outlier Detection of the time spent at POIs: (a) Violin Plot on weekdays, and (b) Violin Plot on weekends.

spent is a crucial input. Previous studies work with assumptions about the time spent. Hence, the developed models enrich the literature with more realistic predictions, where robust regression models are applied to spatial and non-spatial parameters.

The statistical findings reveal a positive beta coefficient for all parameters except for the safety and security levels of the spatial parameters. These outcomes are reasonable because the number of accidents and crimes expresses the safety and security levels. In other words, increasing the number of accidents or crimes decreases the time spent at POIs. However, the lack of a significant effect of safety and security levels on the time spent points that visitors may not be significantly affected by these factors when making decisions about how much time to spend at POIs. As mentioned previously, the explanatory variable with the highest value is relatively the most significant. Hence, the tourist attraction category is the most influential factor that positively increases the time spent at POIs on both weekends and weekdays because tourists tend to spend more time at touristic attractions compared to other types of POIs, regardless of whether it is a weekday or weekend. Generally, the remaining category types (i.e., dining, public facilities, shopping, and entertainment) are the second most important variables, and the rest of the variables have a slight impact on the time spent.

Another evidence on the suitability of the outcomes is the significance of the availability of car parking or PT stations. The availability of car parking or PT stations may affect visitors' decisions about whether to visit a POI and for how long to stay at the location. For example, if a visitor has difficulty finding parking or access to PT stations, they may be less interested to visit a POI or may spend less time at the location. In addition, these results indicate that car owners who are used to direct out-of-vehicle facilities have a lower tolerance for longer walking distances and time delays. This result is expected as illustrated in the spatial analysis results of the availability of car parking and PT stations. The study area is well covered by off-street parking lots (Figure 4). Furthermore, more than 70% of the POIs are within acceptable walking distance, even though bus stations are not considered in this study. Similarly to the findings of [10] and [55], visitors' average rating has a significant impact on the visit duration. In other words, the positive rating (i.e., five stars) means longer time spent at POIs. Generally, the findings of current research are consistent with the outcomes of previous studies [20], which find that spatial factors have a low influence on the visitors' behavior, and conflicts with [21], who encounter this claim.

Obviously, the constant intercept values for the developed models are positively significant. This stipulates that in any case of the dependent variables (i.e., even in the worst case with zero value), there is a time spent at the POIs. In fact, the results are reasonable because some scholars consider time as a commodity as it can provide a utility when operated for specific activities [56]. Individuals' decision-making on the time spent often involves a trade-off between the time spent and other opportunities with the aim of maximizing or minimizing the time spent at POIs according to the preferences. Individuals attempt to raise the amount of time they spend on conducting activities by reducing the number of other opportunities [57]. However, the intercept value for the weekend model is higher than in case of the weekday model. A possible explanation is that people have more free time and are less constrained by work or other commitments on weekends, which may allow them to spend more time at POIs. Weekends are often associated with leisure and recreational activities, which may encourage people to spend more time at POIs compared to weekdays.

The output of current research is important for several studies that involve the time spent, especially those that deal with activity-based models, and examine the optimization of the activity chains. Previous studies, such as [58] and [59], assume the time spent at POIs, while current research work provides realistic values regarding the time spent. Furthermore, the developed models allow the individuals to observe their time spent at touristic, workplace, entrainment, and other destinations, which can be useful for planning journeys and optimizing activity chains.

Furthermore, this study confirms the value of the widespread technologies, such as mobile phones and LBSs, based on recent studies by [10], [60], and [61]. Researchers may obtain a high amount of behavioral data from people by utilizing mobile phone sensors, which has a significant advantage over traditional survey methods. The findings of current research provide insights into visitors' travel behavior, which might be useful for the companies to enhance factors that affect the time spent, especially in case of tourism attractions. In addition, the findings are relevant for transport planners shedding light on impacts of transport variables, such as the availability of car parking lots or PT stations around POIs.

The following limitations of this study have to be highlighted. A sample bias may be present in the popular time data as GPT cannot be used to exactly calculate the real number of visits at specific POIs. However, it is found that GPT is an appropriate source for creating the database because it works with aggregated data. The developed models may not be directly applicable to another study area. This limitation arises primarily due to the variations in spatial parameters across different cities. The spatial parameters considered in the study, i.e., security and safety levels, availability of car parking, and proximity to PT stations, play a crucial role in influencing the time spent at POIs. These parameters are specific to the city of Budapest, which is the focus of the study. Other cities may have different characteristics and factors that impact visitors' behavior. Another limitation is that Google does not provide sociodemographic information about the users, which would be useful to better understand individual behavior. However, this could be a suitable starting point for future study by combining various data sources and applying different data collection techniques. In addition, the combination of GPT with behavioral data on social networking applications, such as Facebook or Instagram, could be applied. Some significant factors, such as weather, PT timetable, and parking policy, should be considered in future studies, as well.

VI. CONCLUSION

Robust regression models are applied to predict the time spent at POIs on weekdays and weekends based on their popularity. To predict the time spent of various POI categories in Budapest, spatial and non-spatial parameters are used. The extracted behavior data from GPT help to understand the relationships of spatial and non-spatial factors and the time spent at POIs. The results reveal that all independent variables significantly affect the time spent based on the regression models on weekdays and weekends except for the number of reviewers. The conclusion of current study shows that the POIs category factor has a substantial impact on the time spent both on weekdays and weekends. In addition, there is a negative relationship between the time spent and the safety or security variables. The results confirm the validity of the developed models. This study makes a contribution by highlighting the value of the GPT data as predictors of behavior insights assisting decision-makers in tracking demand patterns and developing efficient responses to unexpected situations. Additionally, current study provides valuable models, where individuals or tourists can use these models to schedule their activities and optimize their trips. The findings are particularly useful for transportation planners and operators as they shed light on the effect of such variables as parking availability and PT station distances of POIs.

ACKNOWLEDGMENT

The authors are grateful for Wenzhe Sun (Kyoto University), Vishal Mahajan (TU Munich) for creating the Python code and for providing the relevant database. The linguistic revision of current article is prepared by Eszter Tóth.

REFERENCES

[1] T. A. Arentze and H. J. P. Timmermans, "A learning-based transportation oriented simulation system," *Transp. Res. B, Methodol.*, vol. 38, no. 7, pp. 613–633, Aug. 2004, doi: 10.1016/j.trb.2002.10.001.

[2] M. G. McNally and C. R. Rindt, *The Activity-Based Approach*. Bingley, U.K.: Emerald Group Publishing, 2007.

[3] S. J. Choi, S.-B. Park, and K.-Y. Kim, "Estimating category of pois using contextual information," *Indian J. Sci. Technol.*, vol. 8, no. S7, p. 718, Apr. 2015, doi: 10.17485/ijst/2015/v8is7/70475.

[4] K. H. Grantz, H. R. Meredith, D. A. T. Cummings, C. J. E. Metcalf, B. T. Grenfell, J. R. Giles, S. Mehta, S. Solomon, A. Labrique, N. Kishore, C. O. Buckee, and A. Wesolowski, "The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology," *Nature Commun.*, vol. 11, no. 1, pp. 1–8, Sep. 2020, doi: 10.1038/s41467-020-18190-5.

[5] D. Kim, H. Chun, and H. Lee, "Determining the factors that influence college students' adoption of smartphones," *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 3, pp. 578–588, Mar. 2014, doi: 10.1002/asi.22987.

[6] E. Dehaerne, B. Dey, S. Halder, S. D. Gendt, and W. Meert, "Code generation using machine learning: A systematic review," *IEEE Access*, vol. 10, pp. 82434–82455, 2022, doi: 10.1109/ACCESS.2022.3196347.

[7] X. Liu, A. Liu, J. Tadrous, L. He, B. Ji, and Z. Zheng, "IEEE access special section: Intelligent data sensing, collection, and dissemination in mobile computing," *IEEE Access*, vol. 8, pp. 222248–222254, 2020, doi: 10.1109/ACCESS.2020.3040051.

[8] I. A. Junglas and R. T. Watson, "Location-based services," *Commun. ACM*, vol. 51, no. 3, pp. 65–69, 2008, doi: 10.1145/1655925.1656112.

[9] Google. (2020). *Popular Times, Wait Times, and Visit Duration*. [Online]. Available: https://support.google.com/business/answer/6263531?hl=en

[10] M. Möhring, B. Keller, R. Schmidt, and S. Dacko, "Google popular times: Towards a better understanding of tourist customer patronage behavior," *Tourism Rev.*, vol. 76, no. 3, pp. 533–569, May 2021, doi: 10.1108/TR-10-2018-0152.

[11] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, Dec. 2004, doi: 10.1109/TITS.2004.837813.

[12] S. I.-J. Chien and C. M. Kuchipudi, "Dynamic travel time prediction with real-time and historic data," *J. Transp. Eng.*, vol. 129, no. 6, pp. 608–616, 2003, doi: 10.1061/(ASCE)0733-947X(2003)129:6(608).

[13] Y. Duan, V. Yisheng, and F.-Y. Wang, "Travel time prediction with LSTM neural network," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 1053–1058, doi: 10.1109/ITSC.2016.7795686.

[14] J. Ma, J. Chan, G. Ristanoski, S. Rajasegarar, and C. Leckie, "Bus travel time prediction with real-time traffic information," *Transp. Res. C, Emerg. Technol.*, vol. 105, pp. 536–549, Aug. 2019, doi: 10.1016/j.trc.2019.06.008.

[15] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*. Hoboken, NJ, USA: Wiley, 2021.

[16] S. Chakravarty, H. Demirhan, and F. Baser, "Fuzzy regression functions with a noise cluster and the impact of outliers on mainstream machine learning methods in the regression setting," *Appl. Soft Comput.*, vol. 96, Nov. 2020, Art. no. 106535, doi: 10.1016/j.asoc.2020.106535.

[17] A. Carrel, A. Halvorsen, and J. L. Walker, "Passengers' perception of and behavioral adaptation to unreliability in public transportation," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2351, no. 1, pp. 153–162, Jan. 2013, doi: 10.3141/2351-17.

[18] H. Badia, J. Argote-Cabanero, and C. F. Daganzo, "How network structure can boost and shape the demand for bus transit," *Transp. Res. A, Policy Pract.*, vol. 103, pp. 83–94, Sep. 2017, doi: 10.1016/j.tra.2017.05.030.

[19] H. Wang, X. Tang, Y.-H. Kuo, D. Kifer, and Z. Li, "A simple baseline for travel time estimation using large-scale trip data," *ACM Trans. Intell. Syst. Technol. (TIST)*, vol. 10, no. 2, pp. 1–22, 2019.

[20] J. Ma and K. G. Goulias, "Application of Poisson regression models to activity frequency analysis and prediction," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1676, no. 1, pp. 86–94, Jan. 1999, doi: 10.3141/1676-11.

[21] B. Simpson, "Making the connections: Final report on transport and social exclusion. Report by the social exclusion unit 2003," *Proc. Inst. Civil Eng. Municipal Eng.*, vol. 156, no. 2, pp. 81–83, Jun. 2003, doi: 10.1680/muen.2003.156.2.81.

[22] T. Schwanen, "The determinants of shopping duration on workdays in The Netherlands," *J Transp Geogr*, vol. 12, no. 1, pp. 35–48, 2004, doi: 10.1016/S0966-6923(03)00023-1.

[23] B. McKercher and D. Y. Y. Wong, "Understanding tourism behavior: Examining the combined effects of prior visitation history and destination status," *J. Travel Res.*, vol. 43, no. 2, pp. 171–179, Nov. 2004, doi: 10.1177/0047287504268246.

[24] S. W. Litvin and G. H. Kar, "Individualism/collectivism as a moderating factor to the self-image congruity concept," *J. Vacation Marketing*, vol. 10, no. 1, pp. 23–32, Jan. 2004, doi: 10.1177/135676670301000103.

[25] D. Gursoy and K. W. McCleary, "An integrative model of tourists'Information search behavior," *Ann. Tourism Res.*, vol. 31, no. 2, pp. 353–373, Apr. 2004, doi: 10.1016/j.annals.2003.12.004.

[26] B. A. Masberg and L. H. Silverman, "Visitor experiences at heritage sites: A phenomenological approach," *J. Travel Res.*, vol. 34, no. 4, pp. 20–25, Apr. 1996, doi: 10.1177/004728759603400403.

[27] H. J. Kim, D. H. Kim, and J.-H. Chung, "Weekend activity and travel behavior in a developing country: Empirical study using multilevel structural equation models," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1894, no. 1, pp. 99–108, Jan. 2004, doi: 10.3141/1894-11.

[28] B. Dadgostar and R. M. Isotalo, "Factors affecting time spent by near-home tourists in city destinations," *J. Travel Res.*, vol. 31, no. 2, pp. 34–39, Oct. 1992, doi: 10.1177/004728759203100206.

[29] A. Maruyama, N. Shibata, Y. Murata, K. Yasumoto, and M. Ito, "A personal tourism navigation system to support traveling multiple destinations with time restrictions," in *Proc. 18th Int. Conf. Adv. Inf. Netw. Appl. (AINA)*, Mar. 2004, pp. 18–21, doi: 10.1109/AINA.2004.1283747.

[30] C. A. Vassiliadis, C.-V. Priporas, and A. Andronikidis, "An analysis of visitor behaviour using time blocks: A study of ski destinations in Greece," *Tourism Manage.*, vol. 34, pp. 61–70, Feb. 2013, doi: 10.1016/j.tourman.2012.03.013.

[31] S. Yu, C. Wang, K. Liu, and A. Y. Zomaya, "Editorial for IEEE access special section on theoretical foundations for big data applications: Challenges and opportunities," *IEEE Access*, vol. 4, pp. 5730–5732, 2016, doi: 10.1109/ACCESS.2016.2605338.

[32] V. Mahajan, G. Cantelmo, and C. Antoniou, "Explaining demand patterns during COVID-19 using opportunistic data: A case study of the city of Munich," *Eur. Transp. Res. Rev.*, vol. 13, no. 1, pp. 1–14, Dec. 2021, doi: 10.1186/s12544-021-00485-3.

[33] S. G. Stradling, "Transport user needs and marketing public transport," *Proc. Inst. Civil Eng.-Municipal Eng.*, vol. 151, no. 1, pp. 23–28, 2002, doi: 10.1680/muen.151.1.23.38853.

[34] S. Ren and F. Luo, "Research on hourly precipitation preprocessing method based on deep learning," in *Proc. Int. Conf. Digit. Soc. Intell. Syst. (DSInS)*, Dec. 2021, pp. 308–311, doi: 10.1109/DSInS54396.2021.9670603.

[35] L. Huang, Q. Li, and Y. Yue, "Activity identification from GPS trajectories using spatial temporal POIs' attractiveness," in *Proc. 2nd ACM SIGSPATIAL Int. Workshop Location Based Social Netw.*, Nov. 2010, pp. 27–30, doi: 10.1145/1867699.1867704.

[36] E. Celotto, A. Ellero, and P. Ferretti, "Conveying tourist ratings into an overall destination evaluation," *Proc. Social Behav. Sci.*, vol. 188, pp. 35–41, May 2015.

[37] I. Gasco and M. Gattiglio, "Geological map of the upper Gressoney Valley,Western Italian Alps," *J. Maps*, vol. 6, no. 1, pp. 82–102, 2010, doi: 10.4113/jom.2010.1121a.

[38] M. A. Saif, M. M. Zefreh, and A. Torok, "Public transport accessibility: A literature review," *Periodica Polytechnica Transp. Eng.*, vol. 47, no. 1, pp. 36–43, May 2018, doi: 10.3311/PPtr.12072.

[39] A. Mahdi and D. Esztergár-Kiss, "Analysis of the effective factors for hotel selection by using the fuzzy AHP method," *Industry*, vol. 4, vol. 6, no. 2, pp. 79–82, 2021.

[40] Fisherman's Bastion. (2021). *Retrieved 4*. Accessed: Apr. 2021. https://www.google.com/search?q=Fisherman%27s+Bastion&sxsrf

[41] A. Chiarini and F. Brunetti, "What really matters for a successful implementation of lean production? A multiple linear regression model based on European manufacturing companies," *Prod. Planning Control*, vol. 30, no. 13, pp. 1091–1101, Oct. 2019, doi: 10.1080/09537287.2019.1589010.

[42] D. M. Khan, A. Yaqoob, S. Zubair, M. A. Khan, Z. Ahmad, and O. A. Alamri, "Applications of robust regression techniques: An econo- metric approach," *Math. Problems Eng.*, vol. 2021, pp. 1–9, May 2021, doi: 10.1155/2021/6525079.

[43] V. Verardi and C. Croux, "Robust regression in stata," *Stata J., Promoting Commun. Statist. Stata*, vol. 9, no. 3, pp. 439–453, Sep. 2009, doi: 10.1177/1536867x0900900306.

[44] D. Garcia, "Robust smoothing of gridded data in one and higher dimensions with missing values," *Comput. Statist. Data Anal.*, vol. 54, no. 4, pp. 1167–1178, Apr. 2010, doi: 10.1016/j.csda.2009.09.020.

[45] C. Yu and W. Yao, "Robust linear regression: A review and comparison," *Commun. Statist. Simul. Comput.*, vol. 46, no. 8, pp. 6261–6282, Sep. 2017, doi: 10.1080/03610918.2016.1202271.

[46] R. D. Snee, "Validation of regression models: Methods and examples," *Technometrics*, vol. 19, no. 4, pp. 415–428, Nov. 1977, doi: 10.1080/00401706.1977.10489581.

[47] J. H. Steckel and W. R. Vanhonacker, "Cross-validating regression models in marketing research," *Marketing Sci.*, vol. 12, no. 4, pp. 415–427, Nov. 1993, doi: 10.1287/mksc.12.4.415.

[48] K. Al-Sahili and J. Hamadneh, "Establishing parking generation rates/models of selected land uses for Palestinian cities," *Transp. Res. A, Policy Pract.*, vol. 91, pp. 213–222, Sep. 2016, doi: 10.1016/j.tra.2016.06.027.

[49] T. Rakha and R. El Kontar, "Community energy by design: A simulation-based design workflow using measured data clustering to calibrate urban building energy models (UBEMs)," *Environ. Planning B, Urban Anal. City Sci.*, vol. 46, no. 8, pp. 1517–1533, Oct. 2019, doi: 10.1177/2399808319841909.

[50] L. Duncan, "Gazing on communism: Heritage tourism and post communist identities in Germany, Hungary and Romania by light dun can," *Political Nature Cultural Heritage Tourism, Crit. Essays*, vol. 2, no. 2, pp. 269–288, 2017, doi: 10.1080/14616680050027879.

[51] M. K. Smith, I. P. Sziva, and G. Olt, "Overtourism and resident resistance in Budapest," *Tourism Planning Develop.*, vol. 16, no. 4, pp. 376–392, Jul. 2019, doi: 10.1080/21568316.2019.1595705.

[52] A. Mahdi, J. Hamadneh, and D. Esztergár-Kiss, "Modeling of travel behavior in budapest: Leisure travelers," *Transp. Res. Proc.*, vol. 62, pp. 310–317, Jan. 2022, doi: 10.1016/j.trpro.2022.02.039.

[53] T. Hilgert, M. Heilig, M. Kagerbauer, and P. Vortisch, "Modeling week activity schedules for travel demand models," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2666, no. 1, pp. 69–77, Jan. 2017, doi: 10.3141/2666-08.

[54] J. L. Hintze and R. D. Nelson, "Violin plots: A box plot-density trace synergism," *Amer. Statistician*, vol. 52, no. 2, pp. 181–184, May 1998, doi: 10.1080/00031305.1998.10480559.

[55] P. Neirotti, E. Raguseo, and E. Paolucci, "Are customers' reviews creating value in the hospitality industry? Exploring the moderating effects of market positioning," *Int. J. Inf. Manage.*, vol. 36, no. 6, pp. 1133–1143, Dec. 2016, doi: 10.1016/j.ijinfomgt.2016.02.010.

[56] D. A. Hensher, "Valuation of travel time savings," in *A Handbook of Transport Economics*. Cheltenham, U.K.: Edward Elgar, 2011, pp. 135–159, doi: 10.4337/9780857930873.00014.

[57] J. Hamadneh and D. Esztergár-Kiss, "Travel behavior of car travelers with the presence of park-and-ride facilities and autonomous vehicles," *Periodica Polytechnica Transp. Eng.*, vol. 50, no. 1, pp. 101–110, Dec. 2021, doi: 10.3311/PPtr.18020.

[58] D. Esztergár-Kiss, Z. Rózsa, and T. Tettamanti, "Extensions of the activity chain optimization method," *J. Urban Technol.*, vol. 25, no. 2, pp. 125–142, Apr. 2018, doi: 10.1080/10630732.2017.1407998.

[59] D. Rizopoulos and D. Esztergár-Kiss, "A method for the optimization of daily activity chains including electric vehicles," *Energies*, vol. 13, no. 4, p. 906, Feb. 2020, doi: 10.3390/en13040906.

[60] C. Ratti, D. Frenchman, R. M. Pulselli, and S. Williams, "Mobile landscapes: Using location data from cell phones for urban analysis," *Environ. Planning B, Planning Des.*, vol. 33, no. 5, pp. 727–748, Oct. 2006, doi: 10.1068/b32047.

[61] S. Phithakkitnukoon, T. Horanont, A. Witayangkurn, R. Siri, Y. Sekimoto, and R. Shibasaki, "Understanding tourist behavior using large-scale mobile sensing approach: A case study of mobile phone users in Japan," *Pervas. Mobile Comput.*, vol. 18, pp. 18–39, Apr. 2015, doi: 10.1016/j.pmcj.2014.07.003.

ALI JAMAL MAHDI is currently pursuing the Ph.D. degree with the Budapest University of Technology and Economics, with a focus on the realm of tourist behavior. His current research interests include a wide array of disciplines within transportation engineering, including mastery in multicriteria analysis, GIS analysis, optimization, evolutionary algorithms, scheduling, urban planning, statistical analysis, and activity chain analysis.

TAMÁS TETTAMANTI received the M.Sc. and Ph.D. degrees in traffic engineering with the Budapest University of Technology and Economics, Hungary, in 2007 and 2013, respectively. He is currently an Associate Professor with the Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics. He participates in research and industrial projects both as a researcher and a project coordinator. He is the co-author of more than 180 scientific articles, two patents, and several books. His current research interests include road traffic modeling, estimation and control in the cooperative, connected, and automated mobility (CCAM) field, and related co-simulation technology developments. He is a member of the Public Body of the Hungarian Academy of Sciences.

DOMOKOS ESZTERGÁR-KISS was a Fulbright Scholar with the University of California at Davis, in 2021. He is currently a Senior Research Fellow with the Budapest University of Technology and Economics (BME). He is dealing with the development of mobility as a service solution and establishing workplace mobility plans promoting sustainable commuting.