International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 29, No. 5, October, 2024, pp. 1711-1722
Journal homepage: http://iieta.org/journals/isi

Analyzing Road Users' Behavior: A Data Mining Approach Using Google Maps Popular Time and Web Scraping for Rest Area Visitation Patterns on Highways and Toll Roads

Marita Prasetyani[1*], R. Rizal Isnanto[2], Catur Edi Widodo[3]

[1] Doctoral Program of Information System, Universitas Diponegoro, Semarang 50241, Indonesia
[2] Department of Computer Engineering, Universitas Diponegoro, Semarang 50275, Indonesia
[3] Department of Physics, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang 50275, Indonesia

Corresponding Author Email: maritaprasetyani@students.undip.ac.id

https://doi.org/10.18280/isi.290505

Received: 30 March 2024
Revised: 25 September 2024
Accepted: 11 October 2024
Available online: 24 October 2024

Keywords:
rest area, data mining, web scraping, Google Maps, popular time

ABSTRACT

This study aims to analyze the usage patterns of rest areas on highways and toll roads using a data mining approach, utilizing popular time information from Google Maps and web scraping techniques. We collected popular time data for rest areas along highways and expressways using web scraping technology and Google Maps API. The research methodology includes a data collection phase that extracts popular time information from Google Maps using web scraping techniques. The collected data was then analyzed using data mining techniques to identify road user visitation patterns, specifically at rest areas. The analysis included understanding the most popular times visitors visit rest areas and providing deep insights into road user preferences. The findings indicate specific time patterns during which visitors visit rest areas. These findings can have important implications in the planning and managing of road infrastructure, enabling improved services at rest areas according to road users' needs. They can also assist in determining appropriate locations for constructing new rest areas or necessary improvements to existing rest areas.

1. INTRODUCTION

Proper rest times for drivers traveling long distances are essential to avoid fatigue and improve driving safety. Accidents on highways are often caused by the exhaustion of drivers, which various factors can trigger. Eight elements contributing to driver fatigue include exhausting working hours, limited resting time, lack of driving experience, high work pressure, insufficient sleep, algorithmic management, traffic congestion, and excessive workloads [1]. Driver sickness accounts for about 20%–30% of motor vehicle accidents, often resulting in serious injuries or deaths [2]. Traffic accidents are one of the most common causes of death every year [3]. Fatigue is also responsible for around 10%–30% of motor vehicle accidents [4-6].

Rest areas along both toll roads and highways are one of the places provided for resting drivers. Rest areas are essential facilities on toll routes and highways for drivers to avoid fatigue and improve driving safety (Rest Area Design Elements and their Effects on Drivers' Rest Behavior). Rest areas allow drivers or road users to go to small rooms, walk, eat, sleep for a moment, or use a mobile phone. Daytime rest area density ranges from 13% to 17% of total road users [7, 8]. According by Jung et al. [9] rest areas provide more opportunities to avoid sleepy drivers and can reduce the number of accidents by 14%. A rest area can reduce the toll road and highway accident rates [10]. The rest area criteria on tolls and highways include parking, dining and drinking facilities, toilets, information and signs, and security. One of the parameters in rest area planning is the volume of traffic. The existing and projected traffic volumes provide the basis for future resting places, needs, and facilities [11]. Increasing the duration of rest in the rest area positively impacts drowsiness and improves driver performance [12].

The patterns of road users' visits to rest areas are an essential aspect that needs to be understood to improve driving safety and the management of road infrastructure. With the increasing traffic volume and longer travel distances, the need for efficient rest facilities is becoming more urgent. This research aims to fill the gap in the literature by analyzing road users' behavior at rest areas using a data mining approach. By understanding the visit patterns to rest areas, this research's findings are expected to significantly contribute to the planning and management of road infrastructure and enhance road user safety by reducing driver fatigue, which is often a significant cause of traffic accidents.

Much research is focused on rest areas, especially in the Information Systems domain. However, research that deals explicitly with mining visitor data in rest areas has remained small in recent years. Monitoring the number of visitors using GPS tracking and surveys with interview-based questionnaires was done by Wolf et al. [13], but such methods are still very traditional. Geotag photos from the internet and GPS were used to determine visitors' travel. The results by Orsi and Geneletti [14] can be used to estimate the flow of visitors. The combination of GPS Visitor Tracking (GVT) and Recreation

Suitability Mapping (RSM) is used to understand recreational visitor patterns. Data can be inserted into GS planning on various scales. This study can serve as a recommendation to the manager for future development [15]. Korpilo et al. [16] use My Dynamic Forest's web-based Public Participation GIS (PPGIS) device, which combines smartphone GPS tracking, drawing travel routes, and questionnaires to collect visitor information. Data can be used for future planning and management. And predict the number of visitors to German national parks by comparing surveys and photos of social media-based geotags. The results by Sinclair et al. [17] showed quite a representative structure of visitors with an accuracy of 62%-89%. Visitor management has become essential to sustainable destination management in the tourism sector. The size of the 2ha area is relevant to advancing better management and developing it forward [18].

Seeing visitors in two ways: using images from drones and photographs from cameras installed in the area [19]. Monitoring visitor density in the forest using GPS [20]. Passive Mobile Data (PMD) is a data recording from Mobile Network Operators (MNOs) obtained from consumer mobile devices connected to voice networks and public data. The results obtained from such data include the identification of tourists, detection patterns of temporal and spatial distribution, and analysing space and temporal relationships [21]. Urban Green Spaces (UGSs) in Guangzhou, China, surveyed 24 green park points using 595 questionnaires to visitors. The results showed that more visits occurred at night and on weekends, while sweltering days were avoided [22].

Monitor visitors' movements in Parks and Protected Areas (PPAs) using Streetlight, Inc.'s analytical platform of mobile device data. Park management can predict the number of visitors entering the area through which door, the visitor density, and the management and planning of parks and protected areas (PPAs) in the future [23]. Previous research has estimated the number of visitors to a location using radar. Radar is used to measure the length of the vehicle. A vehicle measuring the length could be a bus, so counted many visitors. Small vehicles can be categorized as private cars and contain less. Inappropriate measurement of the vehicle will occur wrong by counting the number of visitors [24]. Stołowe Mts. National Park, Poland, counts the flow of visitors using 39 pyroelectric sensors that are part of the Monitoring System of Tourist Traffic (MSTT). Observations were conducted for 4 years (2017-2019). They identified four types of tourist seasons: in the summer, when the tourists were crowded; in the autumn and spring, when they were still quite crowded; and in the late fall and winter, when they were most secluded. Observations can lead to better management planning and infrastructure improvement [25]. Count the volume of park visits by processing big data from visitors' mobile phones, field observations, and satellite image analysis. Various analyses of such methods are combined to see patterns of visits and build relationships between visitors and the park in implementing cultural events [26]. The Galleria Borghese Museum in Rome, Italy, monitors museum visitors using the Lagrangian IoT tracking device with the Raspberry Pi reception device placed throughout the museum's premises. The data obtained from the device can view visitor behaviour insights and patterns of visitor travel flow within the museum by Centorrino et al. [27].

They use over 20 million geolocation data in tweets from July 2012 to October 2016 about visits to 643 parks in Singapore. The statistical method can reveal the relationship between the park's quality and visitors' patterns in the city park [28]. An event in Oshkosh with 1461 participants conducted a study of the behavioural characteristics of the participants. GPS data is used to obtain a sequence of participants' activities; this sequence characterizes participants' participation in behavioural activities during the event. Machine learning is used to group participants into different segments based on the sequence of activities. Validation is proposed using aerial photos from the estimated crowd and compared with data maps from GPS [29]. GPS-based tracking methodology should be combined with Agent-Based Models (ABM) and GIS techniques. Data from ABM can represent the total number of visitors to EL Capitan Meadow in the summer [30].

In the study by Huang [31], social media platforms are used to track visitors' activities within the National Park. This data is derived from photo geotags and the timing of visitors' posts. It is possible to discern visitor activity patterns, total time spent, and routes taken within the park. However, the accuracy of this data is not fully guaranteed, as not all visitors share their activities on social media. In the Bavarian Forest National Park, Germany, visitor monitoring employs Volunteered Geographic Information (VGI) based on the Global Navigation Satellite System. This data is collected from three platforms: GPSies, Outdooractive, and Komoot, to evaluate pedestrian movements in the park. The most reliable data was found to come from the GPSies and Outdooractive platforms [32].

Several previous studies have examined specific aspects of rest area usage, which monitored the number of visitors using GPS tracking and interview-based surveys. However, the approach used remains quite traditional and does not fully leverage the potential of big data available today. The research by Orsi and Geneletti, which utilized geotagged photos to estimate visitor flow, represents a step forward in understanding visitation patterns, but it remains limited in scope and application.

This study surpasses previous research by employing a more advanced data mining approach, utilizing popular data from Google Maps and web scraping techniques to obtain more accurate and representative information on visit times. It offers a comprehensive analysis of visitor patterns that can inform the development and enhancement of rest area facilities. It contributes new insights to the literature while providing a practical tool for road infrastructure management and road safety policy decision-making. Combining data mining approaches, web scraping techniques, and frequency analysis, the research enhances understanding of road user behavior related to popular times in rest areas, with significant practical implications. A deep understanding of visitor patterns opens opportunities for improved services, such as optimizing facilities during peak hours, enhancing security in frequently visited locations, and developing promotional strategies tailored to road user preferences. These findings can guide future infrastructure improvements and rest area management, delivering tangible benefits in improving service quality and efficiency in the transport sector and supporting the development of new rest areas.

This paper's exhibition is divided into several sections. Section 1 reviews the background, emerging issues, related research, contributions made, and proposed solutions. Section 2 describes the materials, relevant literature, research methods, and routes. Finally, section 4 contains conclusions and opportunities for future research based on this study's findings.

2. MATERIAL AND METHOD

This study follows four distinct stages, as depicted in Figure 1. These stages include (1) Data Sources, (2) Data Mining, (3) Preprocessing, and (4) Analysis. The research framework is described as follows:

Data sources: In this stage, the focus is on obtaining data relevant to visits to rest area locations from Google Maps. Various techniques, such as web scraping or accessing the Google Maps API, can be employed to collect this data.

Data mining: After obtaining the data, the next step is to extract useful information from the collected data. Data mining techniques are applied to identify patterns, trends, and correlations within the data.

Preprocessing: Preprocessing is essential to ensure that the data is in a suitable format for analysis. This phase involves cleaning the data, handling missing values, removing outliers, and modifying variables if necessary.

Analysis: In the final phase, the previously processed data is analyzed to gain insights and draw conclusions. Statistical analysis, visualization techniques, and other analytical methods can be used to interpret the findings from the data.

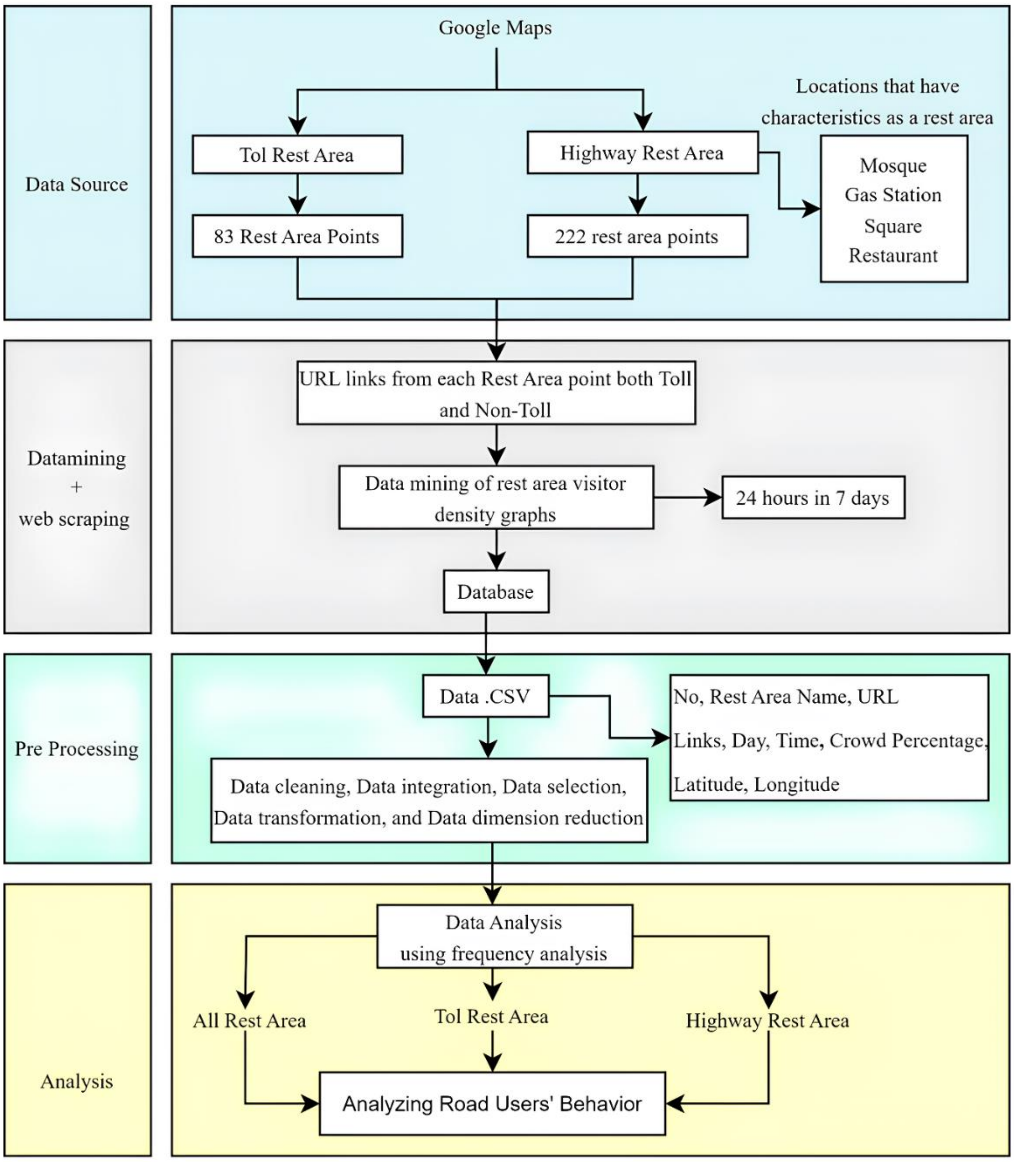


Figure 1. Study framework

2.1 Data source

To ensure that this research can be replicated, we provide detailed steps taken during the data collection process. Popular time data from rest areas were gathered using the Google Maps API, where data was automatically collected daily at specific times to obtain the most accurate information regarding visit patterns. The web scraping process was conducted using tools such as Puppeteer, which allows for the rapid and efficient retrieval of data from the Google Maps website.

Geographic Information System (GIS) is a set of procedures that provide data input, storage, retrieval, mapping, and analysis for spatial data and attributes to support decision-making activities [33]. GIS has grown into a mature field of research and applications involving several academic fields. GIS can support a variety of spatial queries that can be used to support location studies [34]. The widely available Google Maps application exemplifies this GIS capability [35].

Google Maps offers satellite images, road maps, aerial photos, interactive 360 Street View panoramic views, route planning for walking, cars, bikes, and real-time traffic conditions. Google Maps provides satellite images, road maps, aerial photos, interactive 360-degree Street View panoramic views, route planning for walking, driving, biking, and real-time traffic conditions. As part of the Google Maps Platform, the Google Places API offers detailed information about places or locations worldwide, including essential features such as business information, phone numbers, addresses, opening hours, and user reviews. Additionally, Google Maps includes features like travel times and graphs showing average visit times. The Popular Times algorithm is a computational process used by Google to analyze location data from users of the Google Maps application. Based on historical data, this algorithm identifies the busiest and least busy times at a location, with the results presented graphically, which is utilized in this research approach [36]. Traffic conditions or visit data are obtained from Google Maps users on mobile phones, which enables location features. The application then sends the data collected anonymously in real time to Google [37, 38]. The Google Maps service is made interactive, as maps can be moved according to user wishes, such as changing zoom levels and the appearance of the map type [39].

Google Maps is used to obtain old travel information on researches [40-44], view reviews on research [45], estimate traffic speeds [46], and navigate [47].

2.2 Data mining

Data mining is finding patterns and trends that are part of big data. Developments in data mining have several advantages, such as explosive growth in data collection, storage of data on servers so that it is accessible from anywhere and anytime, increased availability of data access from the web and intranets, and tremendous growth in computing power and data storage capacity [48]. Extracts implicit and previously unknown important information from a data set [49, 50]. Data mining can transform decision-making from organizational decisions into data-based decisions [51].

The data mining method obtains the visitor density in the rest area on Google Maps Places by analyzing the historical data of visitors that Google Maps has stored in rest areas of toll roads and highways. The method is implemented using data mining methods of visitor rest area density on Google Maps Place to find out the crowd level or the rest hours of the riders in the rest area.

Web scraping helps extract most information quickly to retrieve, monitor, and facilitate online access to large amounts of information. This method is faster and easier than manually extracting data from the web [52]. The use of web scraping to analyze traffic density has also been conducted by Selmi et al. [53] and Qbouche et al. [54].

Web scraping was performed using Puppeteer, a Node.js library that enables web browsing automation, allowing for systematic data collection and storage in CSV format for further analysis. The data retrieved included visit times and the number of visitors across different rest areas. This automated process, often called web scraping, offers several advantages, such as quickly and efficiently collecting large amounts of information, reducing the risk of human error associated with manual data collection, and enabling real-time data acquisition. Data filtering criteria were applied to ensure that only relevant and accurate data were used in the analysis, including removing incomplete data, anomalies, and data collected during national holidays, which may not be representative. As previously specified, automatic controls were performed using Puppeteer to obtain visitor density data.

2.3 Pre-processing

Data preprocessing was conducted in several stages to ensure data quality before analysis. Initially, the data was automatically saved to the database in CSV format, after which preprocessing was performed. This included data cleaning to remove incomplete, irrelevant, or duplicate data. This step was followed by data integration, selection, and transformation, including numerical data normalization to adjust all variables to a uniform scale, reduce scale variability, and address unit size differences. Outliers that could significantly affect the analysis results were also removed. Additionally, data dimension reduction was applied if necessary. Each preprocessing step, including ensuring that the data type of each column matched the intended analysis, was carefully executed to guarantee that the data used was accurate and representative. Finally, data filtering based on specific criteria was applied, and data integrity was validated to ensure its accuracy and consistency.

2.4 Analysis

Data analysis is the process of applying data mining techniques to frequency analysis to identify trend patterns or information that can provide a deeper understanding of the phenomenon or habits of riders while resting in a rest area. In this context, the pre-processed data explores rider behavior to uncover the most frequently visited times and the most common rest durations. Through this analysis, specific patterns can be identified that reveal riders' preferences for facilities in the rest area, such as a tendency to choose certain times that are either quieter or busier. Additionally, this information can assist authorities in optimizing the management and arrangement of rest areas, for instance, by adjusting services and facilities based on the most frequent visiting times. Thus, this data analysis serves as a tool for understanding rider behavior and as a foundation for strategic decision-making that can enhance comfort and safety in rest areas.

3. RESULT AND DISCUSSION

3.1 Result

The analysis is based on a data mining process that measures visitor density in rest areas, as illustrated in Figure 1. Four processes are involved: data sourcing, data mining, preprocessing, and analysis.

Table 1. Dataset attributes

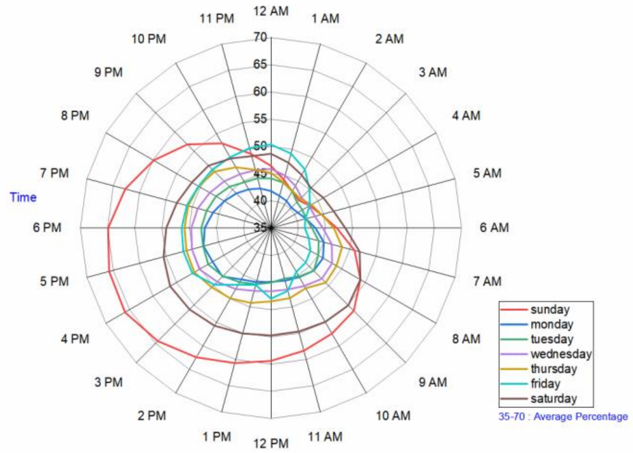
Attribute	Data Type
Rest Area Name	Character
URL Links	Character
Day	Date &Time
Time	Date &Time
Crowd Percentage	Numeric
Latitude	Numeric
Longitude	Numeric

Table 2. Event Log dataset for data mining

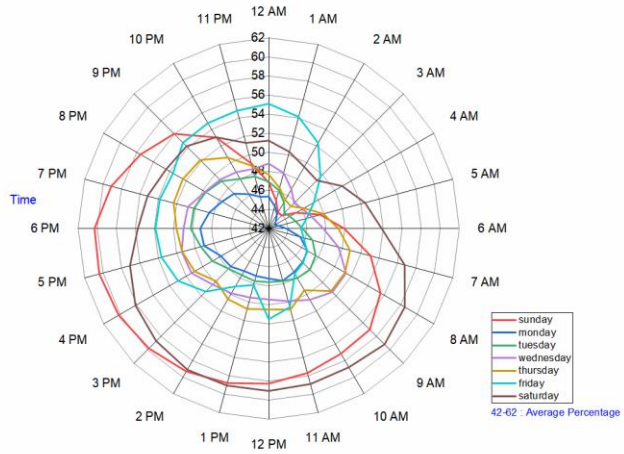
ID	Rest Area Name	URL Links	Day	Time	Crowd Percentage	Latitude	Longitude
1	SPBU Pertamina 34.424.08	https://www.google.com/maps/place/SPBU+Pertamina+34.424.08/@-5.9498502,105.9992657,17.75z/data=!4m5!3m4!1s0x2e4190f0e2e991c7:0x231ecf56896be458!8m2!3d-5.9498428!4d106.0023832	Sunday	4.00	77%	-5.9498428	106.0023832
203	Alun-Alun Kota Cilegon	https://www.google.com/maps/place/Alun-Alun+Kota+Cilegon/@-6.0206315,106.0667692,13.5z/data=!4m5!3m4!1s0x2e418e3d9b577799:0xc02162dd5d691597!8m2!3d-6.009672!4d106.0397286	Monday	22.00	0%	-6.009672	106.0397286
327	Restoran Simpang Raya Merak	https://www.google.com/maps/place/Restoran+Simpang+Raya+Merak/@-5.9373145,106.0026914,16z/data=!4m5!3m4!1s0x2e4190c263124285:0x71f8730af214472b!8m2!3d-5.9357133!4d106.0019939	Monday	12.00	50%	-5.9357133	106.0019939
467	Alun-Alun Kramatwatu	https://www.google.com/maps/place/Alun-Alun+Kramatwatu/@-6.0647087,106.1199234,14z/data=!4m5!3m4!1s0x2e418cf000000001:0x2e30261fc789474a!8m2!3d-6.0647175!4d106.1199106	Sunday	10.00	55%	-6.0647175	106.1199106
1262	Rest Area68A	https://www.google.com/maps/place/Rest+Area68A/@-6.1333205,106.2030988,17z/data=!3m1!4b1!4m5!3m4!1s0x2e41f5a075233417:0xcd80ef1ccb25cd56!8m2!3d-6.1333258!4d106.2052875	Sunday	5.00	16%	-6.1333258	106.2052875
1732	REST AREA PINANG POINT, TOL TANGERANG-JAKARTA KM.14	https://www.google.com/maps/place/REST+AREA+PINANG+POINT,+TOL+TANGERANG-JAKARTA+KM.14/@-6.2139098,106.6754405,17z/data=!3m1!4b1!4m5!3m4!1s0x2e69f9828ca97af7:0x4c99bd509f84be4c!8m2!3d-6.2139151!4d106.6776292	Friday	19.00	99%	-6.2139151	106.6776292
1910	Rest Area KM 38 B Jagorawi	https://www.google.com/maps/place/SPBU+Rest+Area+KM+38+B/@-6.5757825,106.8381783,19z/data=!4m5!3m4!1s0x2e69c5d96d3afd7b:0x1dd25fd6b16e7a34!8m2!3d-6.5764639!4d106.8384964	Saturday	5.00	5%	-6.5764639	106.8384964
2007	Rest Area KM 10A Cibubur Square	https://www.google.com/maps/place/SPBU+Rest+Area+KM+10+A/@-6.340459,106.8897765,18.25z/data=!4m5!3m4!1s0x2e69ecdfb2512a9b:0xfa18605efb69ad23!8m2!3d-6.3401026!4d106.8908905	Wednesday	6.00	33%	-6.3401026	106.8908905
2251	Rest Area KM 21 B	https://www.google.com/maps/place/Rest+Area+KM+21+B/@-6.4352143,106.8908983,17z/data=!3m1!4b1!4m5!3m4!1s0x2e69eac13b2c0835:0x47771e6f0b15b2f4!8m2!3d-6.4352196!4d106.893087	Saturday	10.00	20%	-6.4352196	106.893087
2794	Rest Area KM 42 Jalan Tol Jakarta - Cikampek	https://www.google.com/maps/place/Rest+Area+KM+42+Jalan+Tol+Jakarta+-+Cikampek/@-6.3558648,107.2277162,17z/data=!3m1!4b1!4m5!3m4!1s0x2e699c3a911ea1a7:0x82066b1541bfba0f!8m2!3d-6.3558701!4d107.2299049	Sunday	1.00	63%	-6.3558701	107.2299049

Table 3. Summary of rest area density times (hourly)

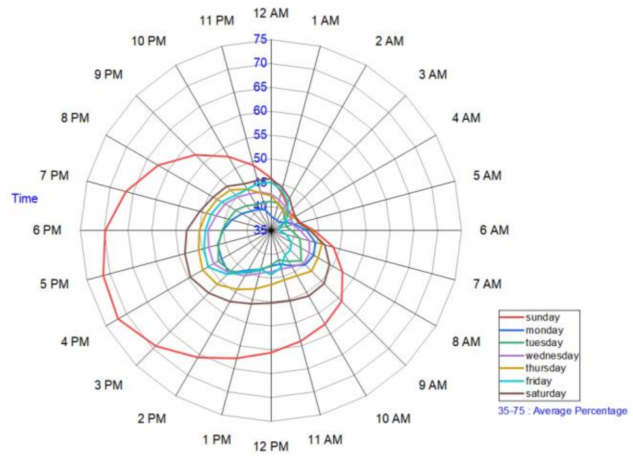
No.	Rest Area	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
a	All rest area	05:00 PM	04:00 PM	04:00 PM	04:00 PM	04:00 PM-05:00 PM 12:00 AM	04:00 PM	04:00 PM-05:00 PM
b	All rest areas from West to East	03:00 PM	04:00 PM	04:00 PM; 08:00 AM	04:00 PM	04:00 PM	04:00 PM	04:00 PM
c	All rest areas from East to West	05:00 PM	06:00 PM	08:00 AM 04:00 PM; 05:00 PM;	08:00 PM	12:00 AM	02:00 PM	05:00 PM
d	Tol	03:00 PM	04:00 PM	04:00 PM	04:00 PM-05:00 PM	05:00 PM	03:00 PM-04:00 PM	05:00 PM
e	Tol from East to West	03:00 PM	04:00 PM	04:00 PM	05:00 PM	05:00 PM	04:00 PM	05:00 PM
f	Tol from West to East	03:00 PM	04:00 PM	04:00 PM	04:00 PM	05:00 PM	08:00 AM	05:00 PM
g	Highway rest area	08:00 AM	04:00 PM	08:00 PM	08:00 PM	04:00 PM; 09:00 PM; 12:00 AM	04:00 PM	04:00 PM
h	Highway rest area from East to West	08:00 AM	04:00 PM	04:00 PM-06:00 PM	04:00 PM	04:00 PM	04:00 PM	04:00 PM
i	Highway rest area from West to East	06:00 PM	07:00 PM	08:00 AM	09:00 PM	09:00 PM	12:00 PM	12:00 PM



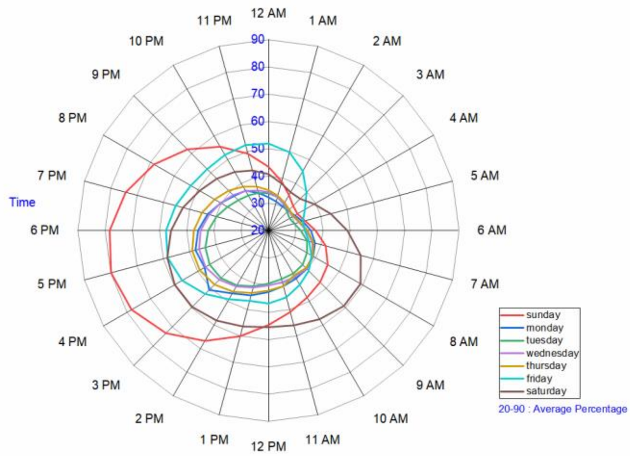
(a) All rest area



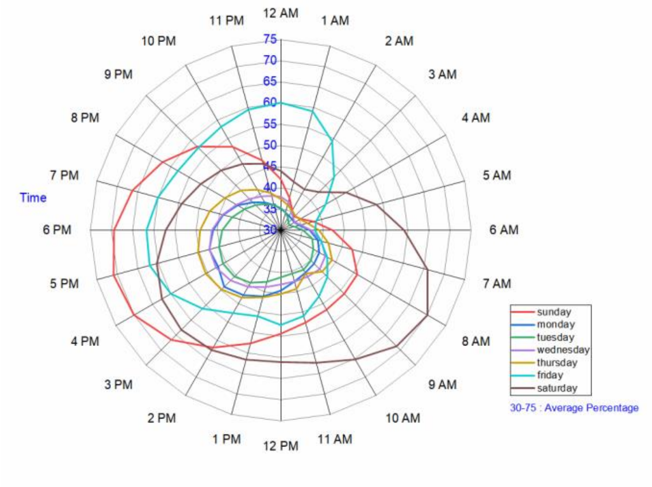
(b) All rest areas West to East



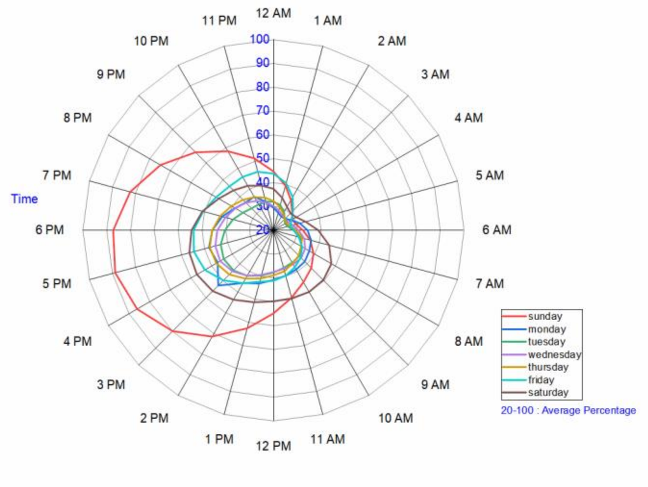
(c) All rest areas East to West



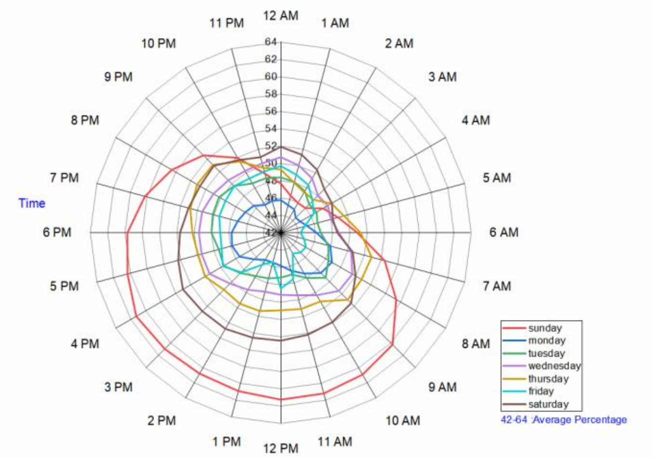
(d) Tol Rest Area



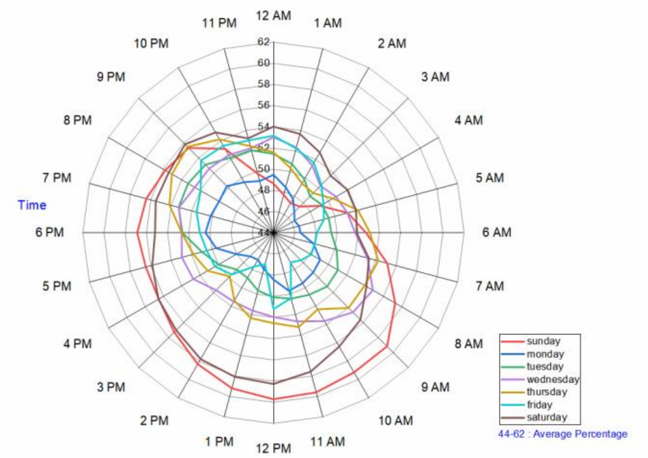
(e) Tol from West to East



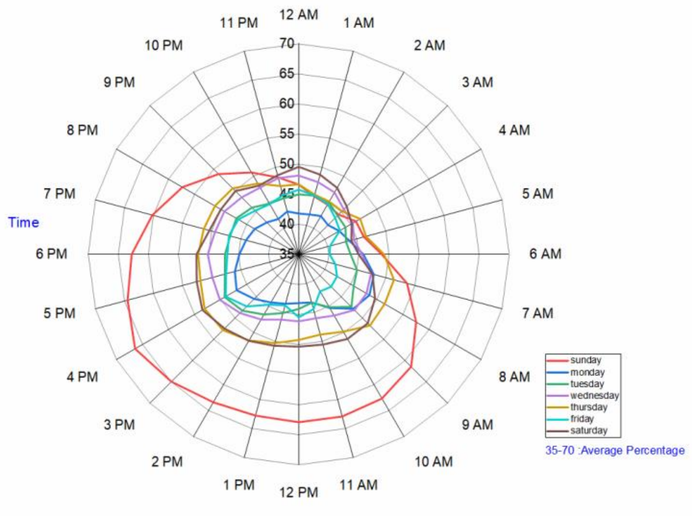
(f) Tol from East to West



(g) Highway rest area



(h) Highway from West to East



(i) Highway from East to West

Figure 2. Visitor density in rest area

The first step involves obtaining a dataset from Google Maps for the northern Java region of Indonesia, with Jakarta as the focal point. Data mining was conducted over a one-week period, from Monday to Sunday, round the clock, at 305 rest area locations. This included 83 rest points in toll areas and 222 in highway areas, yielding a total of 48,676 data points. The data was collected during a period devoid of major or annual holidays. The attributes obtained are listed in Table 1. Logs of the datasets for each rest area are provided in Table 2. An example detailing the information for one of the rest areas can be found in Table 3.

Figure 2 presents the frequency analysis results showing popular time data from Google Maps for all rest areas, including toll and highway rest areas. Use Python libraries Matplotlib to create radar charts. The visitor density across all rest areas, as shown in Figure 2(a). The radar chart illustrates the average percentage of activity across the week, with Sunday (red line) standing out as the day with the highest activity levels, especially between 12 PM and 6 PM. This indicates that Sunday is the peak day for visits or activities. The other days, from Monday through Saturday, show more uniform and lower activity levels, with only slight fluctuations throughout the day. The chart suggests that Sunday experiences the most significant activity, while the rest of the week maintains a more consistent and moderate pattern of activity.

Figure 2(b) illustrates the density across all rest areas from west to east (leaving Jakarta). The radar chart displays the average percentage of activity throughout the week. Sunday (red line) shows the highest activity levels, particularly between 12 PM and 6 PM, making it the peak day for visits or activities. The other days, from Monday to Saturday, exhibit more balanced and moderate activity levels, with slight increases at different times but generally lower than Sunday. The chart indicates that Sunday is the busiest day. At the same time, the rest of the week maintains steadier, less intense activity patterns, highlighting a more even distribution of activity during these days.

The visitor density across all rest areas from east to west (heading to Jakarta), as shown in Figure 2(c). The radar chart illustrates the average percentage of activity across the week. Sunday (red line) shows the highest activity levels, particularly between 12 PM and 6 PM, indicating it as the peak day for visits or activities. The rest of the week, from Monday to Saturday, presents more consistent and moderate activity levels, with slight peaks at various times but generally remaining lower than Sunday. The chart highlights that Sunday experiences the most significant activity. At the same time, the other days exhibit steadier and less intense patterns, suggesting a more balanced distribution of activity throughout the rest of the week.

Visitor density across toll road rest areas from Monday to Thursday. The radar chart depicts the average percentage of activity throughout the week. Sunday (red line) shows the highest activity level, particularly from 12 PM to 6 PM, making it the peak day for visits or activities. The other days, from Monday through Saturday, exhibit more moderate and consistent activity levels, with slight peaks occurring at different times but generally lower than on Sunday. The chart suggests that while Sunday experiences the most significant activity, the rest of the week sees more balanced and subdued activity levels, indicating a consistent but less intense pattern on these days as shown in Figure 2(d).

The visitor density in toll road rest areas from west to east (leaving Jakarta). The radar chart illustrates the average percentage of activity throughout the week, with Sunday (red line) and Monday (green line) showing the most significant activity levels, particularly in the early morning and late afternoon. Sunday has a noticeable rise between 12 PM and 6 PM, indicating it is a peak day for visits or activities. Monday also displays a significant increase in activity, especially during the early hours around 3 AM, which might suggest specific patterns related to late-night or early-morning activity. The other days (Tuesday through Saturday) demonstrate more moderate and consistent levels of activity, with slight variations at different times but generally lower than Sunday and Monday. Overall, the chart suggests that Sunday and Monday are the busiest days, with higher activity levels, while the rest of the week shows more stable and less intense activity patterns as shown in Figure 2(e).

The visitor density in toll road rest areas from east to west (towards Jakarta), as shown in Figure 2(f). The radar chart displays the average percentage of activity across different days of the week, with Sunday (red line) showing a particularly high level of activity, especially between 12 PM and 6 PM. This suggests that Sunday is the peak day for visits or activities, with significant increases in the afternoon and early evening. The other days, from Monday through Saturday, demonstrate much lower and more consistent levels of activity, with minimal variation throughout the day. The chart indicates that Sunday experiences the highest activity levels, while the rest of the week shows subdued and stable activity patterns.

Figure 2(g) the radar chart shows the average percentage of activity throughout the week, with Sunday (red line) standing out as the day with the most significant activity, particularly between 12 PM and 6 PM. This suggests that Sunday is a peak day for visits or activities, with a notable rise in the afternoon and early evening. The other days (Monday through Saturday) exhibit more consistent and moderate levels of activity, with slight variations at different times, but generally lower than Sunday. Overall, the chart indicates that Sunday experiences the highest concentration of activity, while the rest of the week shows more balanced and subdued activity levels.

Figure 2 (h) the visitor density in highway rest areas from west to east (leaving Jakarta). The radar chart illustrates the average percentage of activity across different days of the week, with Sunday (red line) displaying the most significant activity throughout the day, particularly between 12 PM and 6 PM. This indicates that Sunday is a peak day for visits or activities, with a noticeable rise in the afternoon and early evening. The other days of the week, from Monday to Saturday, show more compact and consistent patterns, with activity peaking slightly at different times but overall remaining lower than on Sunday. The chart suggests that the highest concentration of activity occurs on Sunday, while the rest of the week experiences more moderate and evenly distributed levels of activity.

Figure 2(i) The radar chart shows that Sunday (represented by the red line) exhibits the most significant variation in activity throughout the day, with the curve expanding between 12 PM and 7 PM, indicating higher average percentages during the afternoon and evening compared to other days. This suggests that Sunday is a more popular day for visits or activities. In contrast, the lines for Monday through Saturday are more compact and show less variation, with more consistent percentages throughout the day and small peaks at certain times, such as around 6 AM, 4 PM, and 5 PM, indicating increased activity during those periods. Overall, activity levels are consistent from around 5 AM to 7 PM, with higher percentages in the afternoon and evening, particularly on Sunday. The inner areas of the chart, closer to the center, reflect lower percentages, indicating less activity during the night and early morning hours. Sunday stands out with higher and more varied activity levels in the afternoon and evening, suggesting specific factors that drive increased activity or visits on this day compared to other days, which generally show similar patterns with small variations in peak times but overall lower levels of activity than Sunday.

Popular time data from Google Maps has the potential for bias, especially since it only includes users who use the application. This bias can affect the representativeness of the study's findings, particularly in terms of the demographics of visitors who are more likely to use this technology. To address this bias, we recommend using additional data from other sources, such as direct surveys or data from telecommunications service providers, which can offer a more comprehensive view of visitation patterns at rest areas. The analysis in this study primarily relies on 'Popular Times' data from Google Maps. While this data provides valuable insights, it may not fully capture the behavior patterns of all visitors, particularly those who do not use Google Maps. To address potential biases introduced by this method, it is important to acknowledge the limitations of relying solely on this data source. Additionally, integrating supplementary data sources, such as direct surveys, data from telecommunications providers, or other geolocation services, can help mitigate these biases and offer a more comprehensive understanding of visitor behavior at rest areas. This multi-source approach would enhance the representativeness and accuracy of the findings, providing a more holistic view of the visitation patterns.

The increase in visitors during weekends and late afternoons into the evening can be associated with recreational travel patterns, where drivers take longer breaks during leisure trips. This is in contrast to weekdays, where visits tend to be shorter and concentrated at specific times due to time constraints and the pressure to complete journeys.

Additionally, the spikes in visitors on certain days may be influenced by external factors such as weather conditions, special events in the area, or heavier traffic during certain times. These factors should be considered in explaining the findings to provide a more comprehensive understanding of the dynamics of rest area usage.

These findings have significant implications for future policies and practices. By understanding the logic behind these visitation patterns, managers can develop more effective strategies for managing rest area facilities. For instance, increasing parking capacity or enhancing services during peak hours can help alleviate congestion and improve user comfort. Furthermore, these findings can serve as a foundation for better road safety policies, such as adjusting security patrol schedules or adding extra facilities during critical times.

The findings of this study have significant practical implications for rest area management. By identifying peak visitation times, management can optimize resource allocation, such as the placement of security personnel, parking arrangements, and the availability of facilities. Additionally, these findings can be used to design more effective promotional strategies, targeting peak times with special offers or additional services that enhance visitor comfort.

The study results show a consistent visitation pattern across various rest areas, with peak visits occurring late afternoon to evening. This is most likely due to drivers needing to rest after long journeys undertaken during the day. Additionally, the increase in weekend visits reflects a more relaxed recreational travel pattern, where drivers are more likely to take longer rest breaks.

This study enriches the literature by presenting more accurate and applicable findings than previous research. By using popular time data from Google Maps and data mining techniques, we successfully identified visitation patterns that were previously undetectable using traditional methods. This highlights the added value of the approach used in this research.

3.2 Discussion

Popular time analysis results show specific patterns in visitors' visits to rest areas on highways and toll roads. A summary of the visitor density in all rest areas from Monday to Sunday can be seen in Table 3.

A Summary of tabular data presented consistently indicates that the most popular time for visiting the rest area in Google Maps is in the afternoon to evening, from 03.00 pm to 09.00 pm. Few visitors regularly visit the rest area, but a considerable increase occurs on weekends, especially in the morning and afternoon. This pattern suggests that users search for and use rest area services at these hours, possibly due to daytime busyness or as part of a night trip. Using popular time data from Google Maps, we can identify specific periods when rest areas experience a surge in visitor activity. This analysis provides insights into the preferences of road users related to time of visit, which can be used to improve management and service in the rest area.

The novelty of this research lies in its integration of big data from Google Maps with web scraping techniques, providing a detailed analysis of visit patterns that can inform rest area management decisions. Unlike previous studies that focused on static data or small samples, real-time data in this research enhances the accuracy of visitation predictions and provides actionable insights for policymakers. This contribution fills a critical gap in the literature by offering a scalable, data-driven approach to analyzing rest area usage on highways and toll roads, with direct implications for improving road safety and infrastructure efficiency.

The findings of this study have significant practical implications for rest area management and road safety policies. By understanding popular visitation patterns, authorities can optimize resource allocation at rest areas, such as the placement of security personnel, the arrangement of parking facilities, and the availability of amenities like restrooms and dining areas. This also allows managers to design more effective promotional strategies that align with road users' preferences, for instance, by offering additional services during peak hours.

Moreover, this information can be utilized for future infrastructure planning, such as determining the appropriate locations for new rest areas or enhancing existing facilities. A deeper understanding of popular visitation times can also help reduce the risk of traffic accidents related to driver fatigue by providing adequate rest facilities when needed. Therefore, the findings of this study can deliver tangible benefits to improving service quality at rest areas and enhancing efficiency in the transportation sector.

4. CONCLUSIONS

This study demonstrates the potential of using big data analytics, specifically Google Maps' Popular Times data and web scraping techniques, to gain insights into road users' behavior at rest areas. The findings reveal distinct visitation patterns, particularly highlighting the peak Sunday activity between 12 PM and 6 PM. These insights can be leveraged to optimize the management of rest areas by ensuring that resources are allocated efficiently during peak times, enhancing service quality, and improving road safety by mitigating driver fatigue. The study contributes to the academic understanding of rest area usage and provides practical implications for policymakers and infrastructure managers. By utilizing real-time data, this research presents a more accurate and dynamic approach to understanding and responding to the needs of road users. Future research could expand on these findings by incorporating additional data sources to address potential biases and further refine the strategies for rest area development and management.

Building on the findings of this study, future research could explore several avenues to enhance further the understanding and management of rest areas on highways and toll roads. One promising direction is the development of predictive models that incorporate various factors such as weather conditions, traffic flow data, and social and economic variables to forecast visitation patterns more accurately. Additionally, integrating data from diverse sources, such as telecommunications providers or direct surveys, could help mitigate potential biases and provide a more comprehensive view of road user behaviour. Future studies could also focus on real-time monitoring and analysis using advanced data analytics and machine learning techniques to dynamically adjust rest area services and facilities in response to changing conditions. Furthermore, examining the impact of rest area improvements on road safety and driver well-being could offer valuable insights into the effectiveness of these interventions. Overall, the continued integration of emerging technologies and diverse data sources will be crucial in developing more efficient and user-centered rest area management strategies.

REFERENCES

[1] Jaydarifard, S., Behara, K., Baker, D., Paz, A. (2024). Driver fatigue in taxi, ride-hailing, and ridesharing services: A systematic review. Transport Reviews, 44(3): 572-590. https://doi.org/10.1080/01441647.2023.2278446

[2] Garbarino, S., Lino, N., Beelke, M., Carli, F.D., Ferrillo, F. (2001). The contributing role of sleepiness in highway vehicle accidents. Sleep, 24(2): 201-206. https://doi.org/10.1093/sleep/24.2.201

[3] World Health Organization. (2020). Road safety. In Health at a Glance: Asia/Pacific 2020. https://doi.org/10.1787/65afc565-en

[4] Higgins, J.S., Michael, J., Austin, R., Akerstedt, T., Van Dongen, H.P.A., Watson, N., Czeisler, C., Pack, A.I., Rosekind, M.R. (2017). Asleep at the wheel—The road to addressing drowsy driving. Sleep, 40(2): zsx001. https://doi.org/10.1093/sleep/zsx001

[5] Marcus, J.H., Rosekind, M.R. (2017). Fatigue in transportation: NTSB investigations and safety recommendations. Injury Prevention, 23(4): 232-238. https://doi.org/10.1136/injuryprev-2015-041791

[6] Husain, N.A., Mohamad, J., Idris, M.A. (2019). Daily emotional demands on traffic crashes among taxi drivers: Fatigue and safety motivation as mediators. IATSS Research, 43(4): 268-276. https://doi.org/10.1016/j.iatssr.2019.03.001

[7] Al-Kaisy, A., Kirkemo, Z., Veneziano, D., Dorrington, C. (2011). Traffic use of rest areas on rural highways: Recent empirical study. Transportation Research Record, 2255(1): 146-155. https://doi.org/10.3141/2255-16

[8] Pu, D. (1999). Tata Cara Penentuan Lokasi Tempat Istirahat di Jalan Bebas Hambatan. Jakarta.

[9] Jung, S., Joo, S., Oh, C. (2017). Evaluating the effects of supplemental rest areas on freeway crashes caused by drowsy driving. Accident Analysis & Prevention, 99: 356-363. https://doi.org/10.1016/j.aap.2016.12.021

[10] McArthur, A., Kay, J., Savolainen, P.T., Gates, T.J. (2013). Effects of public rest areas on fatigue-related crashes. Transportation Research Record, 2386(1): 16-25. https://doi.org/10.3141/2386-03

[11] Herman, S.C. (1981). Landscape and Environmental Design (Transportation Research 822). National Academy of Sciences.

[12] Cori, J.M., Downey, L.A., Sletten, T.L., et al. (2021). The impact of 7-hour and 11-hour rest breaks between shifts on heavy vehicle truck drivers' sleep, alertness and naturalistic driving performance. Accident Analysis & Prevention, 159: 106224. https://doi.org/10.1016/j.aap.2021.106224

[13] Wolf, I.D., Hagenloh, G., Croft, D.B. (2012). Visitor monitoring along roads and hiking trails: How to determine usage levels in tourist sites. Tourism Management, 33(1): 16-28. https://doi.org/10.1016/j.tourman.2011.01.019

[14] Orsi, F., Geneletti, D. (2013). Using geotagged photographs and GIS analysis to estimate visitor flows in natural areas. Journal for Nature Conservation, 21(5): 359-368. https://doi.org/10.1016/j.jnc.2013.03.001

[15] Beeco, J.A., Hallo, J.C., Brownlee, M.T. (2014). GPS visitor tracking and recreation suitability mapping: Tools for understanding and managing visitor use. Landscape and Urban Planning, 127: 136-145. https://doi.org/10.1016/j.landurbplan.2014.04.002

[16] Korpilo, S., Virtanen, T., Saukkonen, T., Lehvävirta, S. (2018). More than A to B: Understanding and managing visitor spatial behaviour in urban forests using public participation GIS. Journal of Environmental Management, 207: 124-133. https://doi.org/10.1016/j.jenvman.2017.11.020

[17] Sinclair, M., Mayer, M., Woltering, M., Ghermandi, A. (2020). Using social media to estimate visitor provenance and patterns of recreation in Germany's national parks. Journal of Environmental Management, 263: 110418. https://doi.org/10.1016/j.jenvman.2020.110418

[18] Kebete, Y., Wondirad, A. (2019). Visitor management and sustainable destination management nexus in Zegie Peninsula, Northern Ethiopia. Journal of Destination Marketing & Management, 13: 83-98. https://doi.org/10.1016/j.jdmm.2019.03.006

[19] Donaire, J.A., Galí, N., Gulisova, B. (2020). Tracking visitors in crowded spaces using zenith images: Drones and time-lapse. Tourism Management Perspectives, 35: 100680. https://doi.org/10.1016/j.tmp.2020.100680

[20] Pouwels, R., van Eupen, M., Walvoort, D.J., Jochem, R. (2020). Using GPS tracking to understand the impact of management interventions on visitor densities and bird populations. Applied Geography, 116: 102154. https://doi.org/10.1016/j.apgeog.2020.102154

[21] Reif, J., Schmücker, D. (2020). Exploring new ways of visitor tracking using big data sources: Opportunities and limits of passive mobile data for tourism. Journal of Destination Marketing & Management, 18: 100481. https://doi.org/10.1016/j.jdmm.2020.100481

[22] Shan, X.Z. (2020). Association between the time patterns of urban green space visitations and visitor characteristics in a high-density, subtropical city. Cities, 97: 102562. https://doi.org/10.1016/j.cities.2019.102562

[23] Creany, N.E., Monz, C.A., D'Antonio, A., Sisneros-Kidd, A., Wilkins, E.J., Nesbitt, J., Mitrovich, M. (2021). Estimating trail use and visitor spatial distribution using mobile device data: An example from the nature reserve of orange county, California USA. Environmental Challenges, 4: 100171. https://doi.org/10.1016/j.envc.2021.100171

[24] Thórhallsdóttir, G., Ólafsson, R., Jóhannesson, G.T. (2021). A methodology of estimating visitor numbers at an Icelandic destination using a vehicle counter and a radar. Journal of Outdoor Recreation and Tourism, 35: 100378. https://doi.org/10.1016/j.jort.2021.100378

[25] Mateusz, R. (2021). A method to analyze variability and seasonality the visitors in mountain national park in period 2017–2020 (Stołowe Mts. National Park; Poland). Journal of Outdoor Recreation and Tourism, 35: 100407. https://doi.org/10.1016/j.jort.2021.100407

[26] Guan, C., Song, J., Keith, M., Zhang, B., Akiyama, Y., Da, L.J., Shibasaki, R., Sato, T. (2021). Seasonal variations of park visitor volume and park service area in Tokyo: A mixed-method approach combining big data and field observations. Urban Forestry & Urban Greening, 58: 126973. https://doi.org/10.1016/j.ufug.2020.126973

[27] Centorrino, P., Corbetta, A., Cristiani, E., Onofri, E. (2021). Managing crowded museums: Visitors flow measurement, analysis, modeling, and optimization. Journal of Computational Science, 53: 101357. https://doi.org/10.1016/j.jocs.2021.101357

[28] Chuang, I.T., Benita, F., Tunçer, B. (2022). Effects of urban park spatial characteristics on visitor density and diversity: A geolocated social media approach. Landscape and Urban Planning, 226: 104514. https://doi.org/10.1016/j.landurbplan.2022.104514

[29] Abkarian, H., Tahlyan, D., Mahmassani, H., Smilowitz, K. (2022). Characterizing visitor engagement behavior at large-scale events: Activity sequence clustering and ranking using GPS tracking data. Tourism Management, 88: 104421. https://doi.org/10.1016/j.tourman.2021.104421

[30] D'Antonio, A., Monz, C.A., Crabb, B., Baggio, J.A., Howe, P.D. (2022). Proof of concept study using GPS-based tracking data to build agent-based models of visitors' off-trail behavior in nature-based tourism settings. Applied Geography, 147: 102771. https://doi.org/10.1016/j.apgeog.2022.102771

[31] Huang, R. (2023). Analyzing national parks visitor activities using geotagged social media photos. Journal of Environmental Management, 330: 117191. https://doi.org/10.1016/j.jenvman.2022.117191

[32] Horst, L., Taczanowska, K., Porst, F., Arnberger, A. (2023). Evaluation of GNSS-based volunteered geographic information for assessing visitor spatial distribution within protected areas: A case study of the Bavarian Forest national park, Germany. Applied Geography, 150: 102825. https://doi.org/10.1016/j.apgeog.2022.102825

[33] Pick, J.B. (2005). Geographic Information Systems in Business. Idea Group Publishing.

[34] Church, R.L. (2002). Geographical information systems and location science. Computers & Operations Research, 29(6): 541-562. https://doi.org/10.1016/S0305-0548(99)00104-5

[35] Dutkowski, M. (2021). Geographic information science in supporting business decisions. Procedia Computer Science, 192: 2926-2930. https://doi.org/10.1016/j.procs.2021.09.064

[36] Svennerberg, G. (2010). Beginning Google Maps API 3 (2nd ed.). Apress. Retrieved from https://www.amazon.com/Beginning-Google-Maps-API-3/dp/1430228024.

[37] Damar, A.M. (2017). Begini cara google maps pantau arus lalu lintas. https://www.liputan6.com/tekno/read/2825382/begini-cara-google-maps-pantau-arus-lalu-lintas.

[38] Layuk, N.S. (2015). Perancangan aplikasi pendeteksi daerah macet di kota makassar berbasis android. SISITI: Seminar Ilmiah Sistem Informasi dan Teknologi Informasi, 4(1): 214-223.

[39] Antonelli, W. (2021). It Can Take Years for Google Maps to Update Certain Features-Here's How They Get the Data to Update Street View, Traffic, And More. https://www.businessinsider.com/guides/tech/how-often-does-google-maps-update.

[40] Wagner, J. M., Eschbach, M., Vosseberg, K., Gennat, M. (2020). Travel time estimation by means of Google API data. IFAC-PapersOnLine, 53(2): 15434-15439. https://doi.org/10.1016/j.ifacol.2020.12.2365

[41] Fielbaum, A., Jara-Diaz, S. (2021). Assessment of the socio-spatial effects of urban transport investment using Google Maps API. Journal of Transport Geography, 91: 102993. https://doi.org/10.1016/j.jtrangeo.2021.102993

[42] Costa, C., Ha, J., Lee, S. (2021). Spatial disparity of income-weighted accessibility in Brazilian Cities: Application of a Google Maps API. Journal of Transport Geography, 90: 102905. https://doi.org/10.1016/j.jtrangeo.2020.102905

[43] Li, J., Boonaert, J., Doniec, A., Lozenguez, G. (2021). Multi-models machine learning methods for traffic flow estimation from floating car data. Transportation Research Part C: Emerging Technologies, 132: 103389. https://doi.org/10.1016/j.trc.2021.103389

[44] Ali, A., Qureshi, M. A., Shiraz, M., Shamim, A. (2021). Mobile crowd sensing based dynamic traffic efficiency framework for urban traffic congestion control. Sustainable Computing: Informatics and Systems, 32: 100608. https://doi.org/10.1016/j.suscom.2021.100608

[45] Mathayomchan, B., Taecharungroj, V. (2020). "How was your meal?" Examining customer experience using Google maps reviews. International Journal of Hospitality Management, 90: 102641. https://doi.org/10.1016/j.ijhm.2020.102641

[46] Muñoz-Villamizar, A., Solano-Charris, E.L., AzadDisfany, M., Reyes-Rubiano, L. (2021). Study of urban-traffic congestion based on Google Maps API: the case of Boston. IFAC-PapersOnLine, 54(1): 211-216. https://doi.org/10.1016/j.ifacol.2021.08.079

[47] Mohan, V., Kumar, C.G., Yuvaraj, J., Krishnan, A., Amarchand, R., Prabu, R. (2020). Using global positioning system technology and Google My Maps in follow-up studies-An experience from influenza surveillance study, Chennai, India. Spatial and Spatio-temporal Epidemiology, 32: 100321. https://doi.org/10.1016/j.sste.2019.100321

[48] Larose, D.T., Larose, C.D. (2014). Discovering Knowledge in Data: An Introduction to Data Mining (2nd ed.). John Wiley & Sons, Inc.
[49] Witten, I.H., Frank, E., Hall, M.A. (2011). Data Mining: Practical Machine Learning Tools and Techniques (3rd ed.). Elsevier. https://www.sciencedirect.com/book/9780123748560/data-mining-practical-machine-learning-tools-and-techniques.
[50] Han, J., Pei, J. (2012). Data Mining: Concepts and Techniques. Elsevier. https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques.
[51] Li, L., Lin, J., Ouyang, Y., Luo, X.R. (2022). Evaluating the impact of big data analytics usage on the decision-making quality of organizations. Technological Forecasting and Social Change, 175: 121355. https://doi.org/10.1016/j.techfore.2021.121355
[52] Kaur, P. (2022). Sentiment analysis using web scraping for live news data with machine learning algorithms. Materials Today: Proceedings, 65: 3333-3341. https://doi.org/10.1016/j.matpr.2022.05.409
[53] Selmi, A.T.E., Zerarka, M.F., Cheriet, A. (2024). Enhancing K-Means clustering with post-redistribution. Ingénierie des Systèmes d'Information, 29(2): 429-436. https://doi.org/10.18280/isi.290204
[54] Qbouche, K., Rhoulami, K. (2024). Predicting transport mode in the Rabat Region: A machine learning approach. Ingénierie des systèmes d information, 29(2): 493-500. https://doi.org/10.18280/isi.290210