

Big Data in Tourism: Sources, Techniques, and Applications

FARES Imen¹ / Rural Development Study and Research Laboratory, Mohamed El Bachir University Bordj Bou Arréridj, Algeria, imen.fares@univ-bba.dz

BRAHAM CHAOUCH Toufik / Rural Development Study and Research Laboratory, Mohamed El Bachir University Bordj Bou Arréridj, Algeria, toufik.brahamchaouch@univ-bba.dz

Received: 10-04-2026

Accepted: 15-05-2026

Published 08-06-2026

Abstract:

This research provides a comprehensive overview of Big Data sources in tourism, reviews analytical techniques, and explores practical applications for enhancing tourist experiences, destination management, marketing strategies, and decision-making. Adopting an analytical methodology based on academic literature, professional reports, and case studies, the study first defines Big Data and its core characteristics. It then identifies seven major data sources: social media networks, search engines, booking platforms, mobile phone records (CDRs), the Internet of Things (IoT), public and government data, and financial transaction systems. Key findings indicate that Big Data enables a shift from estimation to precision and from monitoring to prediction, facilitating real-time tourist flow tracking, sentiment analysis, demand forecasting, and dynamic pricing. Practical applications include personalized services, congestion management, marketing ROI measurement, and AI-enhanced security. It represents a strategic necessity for sustainable and competitive smart tourism.

Keywords: Big Data, Tourism Development, Smart Tourism.

Jel Classification Codes : Z32 ; C55 ; D12.

Résumé :

Cette recherche propose un aperçu complet des sources de données massives (Big Data) dans le tourisme, examine les techniques analytiques et explore les applications pratiques pour améliorer les expériences touristiques, la gestion des destinations, les stratégies marketing et la prise de décision. Adoptant une méthodologie analytique fondée sur la littérature académique, les rapports professionnels et les études de cas, l'étude définit d'abord le Big Data et ses caractéristiques fondamentales. Elle identifie ensuite sept sources majeures de données : les réseaux sociaux, les moteurs de recherche, les plateformes de réservation, les enregistrements d'appels mobiles (CDR), l'Internet des objets (IoT), les données publiques et gouvernementales, ainsi que les systèmes de transactions financières. Les principales conclusions indiquent que le Big Data permet un passage de l'estimation à la précision et de la surveillance à la prédiction, facilitant le suivi en temps réel des flux touristiques, l'analyse des sentiments, la prévision de la demande et la tarification dynamique. Les applications pratiques incluent les services personnalisés, la gestion de la congestion, la mesure du retour sur investissement marketing et la sécurité renforcée par l'IA. Il représente une nécessité stratégique pour un tourisme intelligent, durable et compétitif.

Mots clés : Big Data, Développement touristique, Smart Tourisme.

Jel Classification Codes : L53 ; L78 ; M00.

¹. Corresponding author: FARES Imen, imen.fares@univ-bba.dz



I. Introduction

The rapid development of media and communication technologies—especially mobile and smart devices over the past two decades—along with the widespread use of the Internet, has radically changed how information is produced, transmitted, and used. Digital platforms and social networks, alongside traditional information sources, have become open spaces where massive amounts of data flow daily from human interactions and visual or textual contributions. This data ranges from a fleeting tweet to remote sensing data in science, and from consumer preferences in e-markets to people's cultural and social behaviors.

In this context, the concept of Big Data has emerged as one of the most prominent manifestations of this digital revolution. Data is no longer merely administrative records or periodic statistics; it has become a key driver of creativity, innovation, value creation, and evidence-based decision-making across various economic activities.

The tourism sector is among the most affected by this digital transformation. Given its intangible service nature, information related to tourism offerings—and how they are presented—is the main driver of purchase decisions. Massive amounts of data are generated from the interactions of tourists, service institutions, and public authorities at every stage of a trip: from destination searches on search engines, to electronic bookings on travel platforms, to sharing photos and experiences on social media, not to mention data collected indirectly via mobile phones and smart sensors. This data has become a real indicator for understanding tourist behavior, predicting demand trends, improving service quality, and managing tourist destinations intelligently and sustainably.

Accordingly, this research aims to provide a comprehensive overview of Big Data sources in the tourism sector, review the various techniques and methodologies used in its analysis, and focus on its practical applications in improving tourist experience, developing marketing strategies, enhancing destination management efficiency, and supporting decision-making. This will be achieved through an analytical methodology based on a review of academic literature, professional reports, and leading case studies.

II. Definition of Big Data:

Definitions of Big Data vary across organizations and research centers. These differences stem from each party's nature, objectives, technical capabilities, organizational resources, strategic goals, and the type of data available. Government agencies emphasize Big Data's role in supporting public policies, improving public sector efficiency, and advancing societies toward sustainable development goals.

For example, according to the European Union, Big Data refers to massive amounts of data produced at high speed from diverse sources—by people or machines (e.g., climate sensors, satellite images, digital photos, transaction records, GPS signals). When effectively analyzed using appropriate technologies, this data can benefit traditional sectors like transportation, health, and manufacturing. It increases productivity, improves business intelligence, generates innovative information products and services, reduces costs, enhances public sector efficiency, and helps address societal challenges (European Union , 2025).

The United Nations, through its Department of Economic and Social Affairs, focuses on Big Data's contribution to modernizing official statistical systems to keep pace with the digital age. It views Big Data as a complementary factor to traditional statistics—not a substitute—in tracking the Sustainable Development Goals. This is achieved by improving the quality and accuracy of current statistical estimates and generating new indicators and information that were not previously possible, thereby providing new analytical angles across different statistical domains (Department of Economic and Social Affairs, 2022).

Similarly, the Organization for Economic Co-operation and Development (OECD) views Big Data as vast flows of data produced through information and communication technologies—including the Internet—as well as ubiquitous wired sensors that monitor physical world activities. Real-time analysis of this data helps drive knowledge, create value throughout society, enhance new products and markets, stimulate entirely new business models, and transform most sectors in OECD countries, thereby enhancing economic competitiveness and productivity growth (OCDE, 2015).

Business companies, in contrast, focus their definitions on the large volume and immense diversity of Big Data, which require advanced operational techniques to support managerial decision-making and enhance corporate competitiveness. The International Telecommunication Union (ITU) describes Big Data as massive and complex datasets made available through digitalization that cannot be processed using traditional techniques. To extract meaningful information, Big Data analytics uses advanced computing techniques to identify patterns, trends, and repetitions within datasets (ITU, 2021).

The research institution TechAmerica, in its pioneering 2012 report "Demystifying Big Data: A Practical Guide to Transforming the Business of Government," states that "big data is large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information" (ESCWA, 2021).

Meanwhile, IBM focuses on advanced technologies and sophisticated analysis methods such as artificial intelligence and machine learning. Due to its massive volume and complex nature—which traditional data management systems cannot handle—Big Data requires distributed processing systems for proper collection, management, and analysis (Badman & Kosinski, 2024).

Oracle, in contrast, views Big Data as extremely large and complex datasets that are difficult to manage or analyze using traditional tools. Thanks to this growing volume, companies can now make more accurate and effective business decisions. However, maximizing benefit from Big Data is not limited to analysis alone; it is a comprehensive exploration process that requires insightful analysts, business users, and executives capable of asking the right questions, identifying patterns, making informed assumptions, and predicting behavior (Chen & Strategist, 2024).

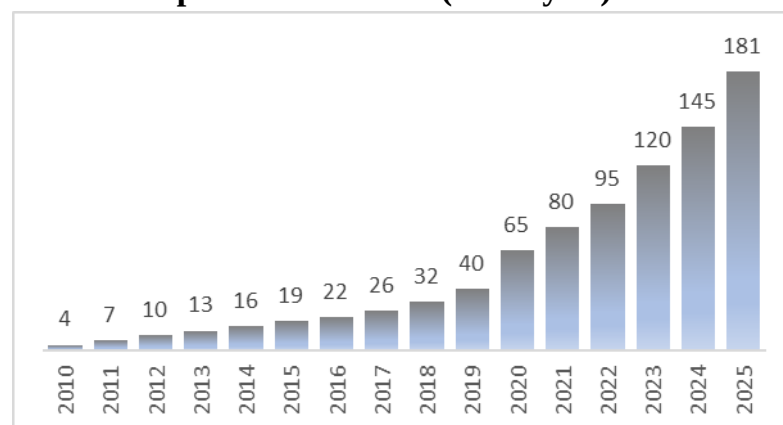
III. Characteristics of Big Data:

Big Data possesses a set of characteristics that impose specific methodologies and techniques for collection, storage, and processing. These are known as the 5 V's, summarized below.

III.1 Volume:

Volume refers to the massive amount of data currently being collected and stored, which is characterized by continuous growth. This volume cannot be collected, stored, or processed using traditional computing and statistical methods, necessitating more advanced techniques.

Figure (1): Volume of digital data created and replicated worldwide during the period 2010-2025 (zettabytes)



Source: (Moghaddam, 2024)

As shown in Figure 1 global internet data created and replicated grew from 4 zettabytes in 2010 to 65 zettabytes by 2025. Daily production reached approximately 328.77 exabytes in 2023. The total is expected to hit 145 ZB in

2024 and 181 ZB in 2025—a 2.8-fold increase in just five years (Moghaddam, 2024)

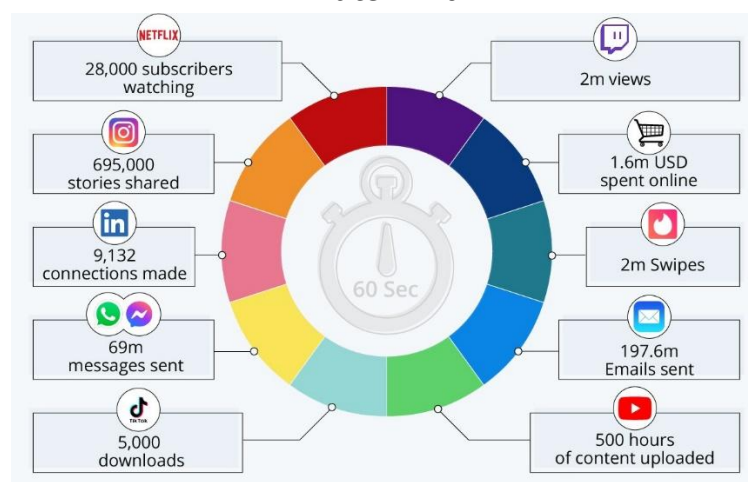
III.2 Velocity:

Big Data is not only characterized by its volume but also by the rapid pace at which data is generated due to its continuous flow from various sources. This results from the widespread use of the Internet in personal activities—via smartphones and various data exchanges (emails, videos, social media, etc.)—as well as its application in finance, industry, and scientific research.

The "Digital 2026 Global Overview Report" issued by DataReportal confirms that the digital world reached significant milestones in 2026. The number of internet users exceeded 6 billion, representing 73.2% of the world's population, while social media user identities rose to 5.66 billion, forming a supermajority of the planet's population. The average user spends more than 2.5 hours per day browsing social media and video platforms. Globally, advertising spending is expected to exceed \$1.16 trillion in 2025, with digital channels dominating 74.4% of this amount, driven by growth in search ads (11.1%) and social media ads (13.6%). The importance of velocity is evident in time-sensitive sectors such as financial markets and healthcare (DataReportal, 2025).

According to data compiled by Lori Lewis and published on AllAccess (via Statista), 60 seconds on the web in 2021 included more than 500 hours of content uploaded to YouTube, 695,000 stories shared on Instagram, nearly 70 million messages sent via WhatsApp and Facebook Messenger, in addition to \$1.6 million spent online (Jenik, 2021), as illustrated in Figure 2.

Figure (2): Estimate of the volume of data circulated on the internet in one minute in 2021



Source: (Jenik, 2021)



III.3 Variety:

Big Data is characterized by great diversity and heterogeneity. We distinguish three main types of data:

- **Structured Data:** This follows a clear format with defined relationships between fields, stored in traditional or modern databases (e.g., sales invoices, financial records).
- **Semi-structured Data:** This contains tags or markers that identify structure and content but does not require a rigid schema. It is flexible, scalable, and can be converted into tabular format fairly easily. Examples include server log files (clicks, browsing), smartwatch sensor data (heart rate, steps), weather data, and industrial production line sensors.
- **Unstructured Data:** Constituting 80–90% of all digital data, this is the largest and fastest-growing type. It does not follow a pre-defined model or structure and cannot be stored in traditional rows and columns. It requires advanced techniques such as artificial intelligence and natural language processing to extract value. Examples include email messages, social media posts, news articles, technical support chat logs, images (medical scans, satellite images), audio recordings (customer service calls), and video recordings (surveillance cameras).

III.4 Veracity

Among the main challenges posed by Big Data is not only its volume and speed but also the credibility, reliability, and accuracy of the data, especially as it comes from diverse sources. This poses a real challenge to data quality and its reliability for making critical decisions. Unreliable Big Data is not just worthless—it can be misleading and dangerous, potentially causing severe economic losses. Therefore, advanced techniques and sophisticated mechanisms must be adopted to review and clean data before use, especially in sensitive analyses.

III.5 Value

Value is the central purpose of Big Data's other characteristics (volume, velocity, variety, veracity). It refers to whether data enables sound, timely decisions—helping organizations understand customers and markets, extract actionable insights, and gain a long-term competitive advantage.

Figure (03): The five main characteristics of Big Data (The 5 Vs)



Source : Created by the authors

IV. Sources of Big Data in Tourism Activities:

Tourism activities rank among the most data-intensive sectors, generating vast amounts of Big Data at every stage—from destination searches on search engines and electronic bookings via travel platforms, to sharing photos and experiences on social media, not to mention data collected indirectly through mobile phones and smart sensors. Therefore, it is essential to provide a comprehensive overview of Big Data sources in the tourism sector, review the techniques and methodologies used to analyze them, and explore how they can be leveraged to develop tourism activity.

IV.1 Social Media Networks

Social media networks are a major source of Big Data in tourism. Tourists frequently use these platforms during their travels, aided by lightweight, easy-to-use mobile devices like smartphones and tablets. User interactions—such as posting photos, comments, hashtags, likes, and reviews—automatically generate large-scale information and data.

Posting photos, videos, or adding a specific location instantly generates data about individuals' coordinates, allowing the construction of so-called heat maps. These maps provide a real picture of the most visited landmarks, preferred visitor routes, and crowding intensity at various facilities.

Moreover, the tweets, comments, and textual posts that tourists publish to express their experiences—whether related to hotels, restaurants, or other services—represent a treasure trove of information about tourists' experiences, feelings, preferences, and complaints. These texts are processed using specialized techniques and carefully designed algorithms that automatically extract keywords



and classify them on a scale (positive, negative, neutral), then organize them by topic (accommodation, food, cleanliness, services, etc.). This helps identify the most in-demand hotel specifications or services that generate complaints or praise, thereby improving services and analyzing competitors. It also enables personalized recommendations based on content (photos, videos, comment nature) or user profile, where digital fingerprints are created for users with similar behavior to offer relevant deals.

A 2025 study analyzed the economic dimensions of tourism destination image using an experimental methodology. The researchers used a sample of 2,577 TripAdvisor reviews for Kastoria, Greece, written in 17 languages by 1,631 tourists from 45 countries, along with 2,598 attached photos. Results revealed clear economic concentration on two main landmarks: Lake Kastoria (33.1% of reviews) and the Dragon Cave (29.2%), which reflected revenue distribution across tourism sectors. Temporal analysis showed sharp seasonality in demand, with price elasticity being important for dynamic pricing strategies. The study also found a statistical correlation between tourist country of origin (based on review language) and spending patterns. Analysis of photos without people indicated that the platform is used to evaluate attraction quality rather than for self-promotion, impacting digital marketing strategies. Review processing revealed positive sentiments among tourists, which can be economically translated into higher willingness to pay and repeat visitation (Yannacopoulou & Kallinikos, 2025).

IV.2 Search Engines:

While social media platforms provide valuable insights into past travel experiences, search engines offer a different but complementary perspective by revealing users' future intentions and desired travel experiences. Search engines like Google serve as genuine indicators of short-term travel plans—both destinations and types of tourism products users are interested in acquiring.

In-depth analysis of this data using models such as Natural Language Processing (NLP) greatly facilitates predicting tourist arrivals well before they occur, as well as the time gap between search date and actual arrival. This guides proactive planning of supply and staffing within tourism structures, while also helping adjust offerings by introducing new products and services that align with tourist aspirations and address potential concerns such as security and cleanliness.

A study in Australia used Google search engine data to understand tourism demand dynamics as part of evaluating digital marketing campaign effectiveness. Analysis of search query data from 2005 to 2021 showed that a positive shock in online search intensity explains more than 35% of the variation in international arrivals and more than 50% of the variation in overnight stays. The impact of

search intensity was stronger for international tourists compared to domestic ones, and a 10% increase in the search intensity index led to a \$1.4 million increase in tourism GDP. These results have enabled Australian authorities to use search behavior as a real-time metric to evaluate digital campaigns (such as "Come and Say G'day"), plan efficient resource allocation, and gauge tourist sentiment during shocks such as pandemics (Duncan & Hailemariam, 2025).

In Singapore, researchers developed a forecasting model integrating Google Trends search data with other external factors to predict tourism demand in the post-COVID-19 period. They applied a systematic selection process to identify the most effective subset of external variables. Results showed that the proposed model outperformed traditional models (SARIMA, Holt-Winters, Prophet) and machine learning models (LSTM, RNN), achieving the lowest Mean Absolute Percentage Error (MAPE) of 7.32% when forecasting over 24 months (2023-2024). This enables decision-makers to use the model for strategic planning and resource allocation with greater precision (Lee, 2025).

IV.3 Booking and Travel Platforms:

Modern information and communication technologies—particularly the internet—have enabled a significant leap forward in managing and organizing services within tourism establishments. These include:

- Electronic booking and payment (locally and internationally), allowing automated, real-time accounting processing.
- Digital presentation of products and information related to tourism activities (destination, date, time, service components), with the ability to instantly inform customers of any sudden changes.
- Real-time transmission of vast amounts of information between operators, suppliers, and customers (email, audio, video).
- Online customer forms that can be processed automatically to enhance or personalize services.
- Real-time inventory of all sales and facility utilization status, thereby reducing idle capacity.

These technologies have reduced costs and improved productivity and efficiency within tourism establishments. Key factors enabling these benefits include digitization of travel-related documents and files, as well as automation in service tracking and file processing, which reduces processing times and minimizes errors. Furthermore, information sharing and collaboration among multiple operators make data rich and inclusive of various customer needs while reducing the unit cost of information processing and expanding opportunities for updating data when necessary.



Recent data from the French National Center for Tourism Resources indicate that the travel and tourism sector ranks as the leading e-commerce sector in France in terms of buyers. Approximately 6 out of every 10 internet users (59.3%) made at least one purchase via travel websites during the twelve months prior to the study, achieving 15% growth compared to the previous year and bringing total market value to €60 billion. This growth is attributed to increasing smartphone reliance and the expansion of online booking platforms offering competitive deals. Accommodation services topped online travel purchases, with 45% of buyers booking accommodation online. Booking.com leads this field with 36% of total buyers, followed by Airbnb with 30%. For transport services, 29% of buyers purchased tickets (air, train, bus) online, with SNCF Connect and Air France dominating this sector (Nadiras, 2024).

A study published by the French Federation of E-commerce and Distance Selling (FEVAD) indicates that in 2023, the tourism and travel sector became the leading e-commerce sector in terms of customers, surpassing fashion, technical products, and household appliances. This reflects consumers' tendency to spend budgets on services—particularly leisure activities—and their preference for digital channels, with 41% preferring to book trips online rather than visiting physical agencies. Younger age groups are most influenced by this trend: the 25-34 age group is the most active, using an average of 3.5 different websites to book travel, while the 15-24 age group is most inclined to purchase transport services through this channel (41% of them), especially train and bus tickets (FEVAD, 2024).

This dynamism in online booking platforms has become a continuous and rapid generator of data, both in quantity and quality. Through precise analysis using specialized algorithms, deep learning, machine learning, and other new mechanisms, a new vision for managing tourism activities becomes possible. This improves services and enhances competitive position by automatically tracking reviews and comments—especially negative ones related to services and facility quality. Strengths and weaknesses can be precisely identified across different aspects of tourism service (cleanliness, location, price, ancillary services, etc.), going beyond general assessments, all within a secure framework that detects fake and misleading reviews. This can be leveraged to improve customer experiences, provide the most attractive products, and offer personalized services based on individual preferences, browsing behavior, or categorical segmentation (purpose of stay, family trips, age groups, etc.).

Moreover, these platforms represent vast databases linked to booking dates, departure dates, cancellation patterns, competitors' prices, and amounts tourists have paid. These data allow constructing models to understand complex relationships between factors, predict actual tourism demand, and determine

occupancy rates. This enables better management of staffing levels and optimal resource allocation, achieving cost savings. Furthermore, real-time monitoring of demand levels and competitors' prices allows adopting dynamic pricing policies that automatically adapt to demand in time and space to maximize revenues—known as Yield Management.

IV.4 Mobile Phones and the Internet of Things:

Mobile phones are now widespread and continuously evolving. Their technological development and diversified services have enabled their use across various fields of tourism activity. In addition to allowing tourists to check email, read news and research, and communicate with family and friends, mobile phones serve as reliable information sources during trips. They provide explanations and comments about visited areas and attractions, enriching user experience beyond paper maps and diagrams, which may contain outdated information or insufficient space. Furthermore, tourists are not obliged to join guided tour groups, as these devices free them from constraints related to preset timetables and routes defined by tour guides.

However, in parallel, these devices generate massive amounts of data that can be used to monitor tourism activity. Every interaction between a user's phone and the communication network (calls, SMS, mobile data) is recorded in detailed logs called Call Detail Records (CDRs), which are automatically generated by mobile network operators. These records capture information about the transaction, including its nature, timing, duration, user number, and the cell tower used.

When a tourist travels from a foreign country to a destination, their phone automatically registers with the visited country's network. These records can then be analyzed to extract information about the number of tourists who visited the area (by counting foreign phones connected based on country of origin using international dialing codes), as well as length of stay (by calculating the difference between first and last connection to the local network). Mobility patterns can be analyzed using the sequence of cell towers to which the tourist's phone connects, allowing mapping of travel routes, identification of most visited destinations (towers attracting the highest number of foreign phones), and most frequently used routes between attractions. This enables prediction of congestion points and peak times by monitoring real-time changes in visitor numbers based on time of day or season.

The Slovenian National Bureau of Statistics and the Bank of Slovenia used telecommunications operator data to measure internal and external tourism flows, aiming to provide more frequent and accurate tourism data. The methodology classified roaming users into four categories based on network connection



duration over two consecutive days: Tourist (18-48 hours), Same-day visitor (12-18 hours), Half-day visitor (6-12 hours), and Transit passenger (less than 6 hours). Data is collected monthly from all network operators (excluding those with market shares below 0.5%) via the telecommunications regulatory authority, and experimental quarterly results are published two months after the reference period.

Results demonstrated the ability of this data to provide wide geographical coverage (more than 100 countries compared to 6-10 countries in traditional surveys), as well as to analyze the impact of major events such as the COVID-19 pandemic or Olympic Games on tourism flows. Despite gaps and challenges—potential overestimation (counting truck drivers, daily workers, or smart devices not associated with tourists) or underestimation (excluding those without mobile phones)—which necessitate integrating this data with other sources and continuing traditional surveys for qualitative information, this experience confirms that mobile network operator data constitutes a powerful complementary tool for monitoring tourism flow trends more frequently and in greater detail (OECD, 2025).

Beyond mobile phones, data generated by the Internet of Things (IoT) is also valuable. Although many pre-existing sensors operated within closed, isolated systems, with data primarily used for direct monitoring and local surveillance, IoT has become a Big Data source as we aggregate data from thousands or millions of devices, store it, and analyze it at scale to discover previously invisible patterns and insights. Examples include smart electricity and water meter data, factory sensor data (temperature, pressure, humidity), precision sensors, and surveillance equipment (high-resolution cameras, thermal cameras).

Tourism activities have not been immune to this development. IoT has enabled a qualitative leap in tourism management and development. Tourist sites can now be transformed into smart entities that automatically interact with variables related to site management and operation. Interactive screens monitor visitor crowd flow via intelligent routing to reduce congestion and avoid overcrowding in facilities (parking lots, restaurants, entertainment facilities). Surveillance cameras managed by advanced systems allow sentiment analysis and abnormal behavior detection, enhancing visitor safety and security. Various sensors measure and monitor public indicators such as energy consumption, climatic conditions, and air quality, providing distinctive tourist experiences.

The management experience of the archaeological site of Pompeii in Italy—spanning 66 hectares and receiving over four million visitors annually—illustrates this well. The site adopted the Integrated Multidisciplinary Model for Safety and Security Management (IMMSSM), an integrated technological system based on the Internet of Things (IoT/IoE) aimed at enhancing safety and security,

managing the site, and improving visitor experience. This was achieved by establishing a communications infrastructure based on wired and wireless connections, AI-powered video surveillance systems, drones, environmental and structural sensors, and satellite data to monitor the city perimeter, entrances, passageways, open spaces, and gathering points. AI-powered cameras were deployed to detect anomalous behaviors (intrusions, vandalism, abandoned objects), alongside the SMI system, which provides a digital twin of the site and analyzes sentiment from social media.

Performance results showed availability of 99.99%, intelligent analysis accuracy of 98.8%, and overall response time of 92 milliseconds. This system represents the first large-scale application of its kind at an archaeological site and offers a reference model adaptable to other cultural heritage sites worldwide (Bruni & Garzia, 2025).

IV.5 Public and Government Data:

Public and government data refers to information produced by governmental bodies through their statistical agencies, focusing on demographic and economic statistics such as income, health, education, and transportation. Examples include the official open data portals of the United States (data.gov) and the United Kingdom (data.gov.uk). International organizations such as the World Bank, World Health Organization, and European Union also publish specialized information according to their fields of activity.

This type of data has wide-ranging applications in tourism management at both macro and local levels. For instance, data produced by authorities managing border crossings—using electronic management systems and biometric registration—can be used not only for border control and security risk identification but also for real-time counting of arriving and departing tourists, as well as classifying them into homogeneous groups according to objective criteria (nationality, country of origin, entry and exit points, age and gender, actual length of stay, etc.).

The experience of British authorities in 2014 is worth mentioning. Under the e-Borders program, transport and airline companies were required to transmit and collect all passenger information and booking data electronically between 24 hours and 30 minutes before travel. This data was to be used for security border monitoring, with the process carried out within the airline's operational system by providing a customer interface for passengers to fill in this data.

Despite the additional cost incurred by airlines in updating technical systems to accommodate this additional data volume and increased transaction time, these companies leveraged this process—originally a compliance requirement—and turned it into part of the added value offered to tourists. It



represents a new point of interaction with customers to offer additional services (travel insurance, hotel bookings, car rentals, etc.), as well as personalized commercial offers based on provided data (age, gender, destination, etc.). Furthermore, this data helps organize and speed up airport check-in procedures, reducing waiting time—a commercial advantage that increases customer satisfaction and loyalty, transforming mandatory administrative procedures into digital experiences that enhance commercial tourism activity (Dibb, Ball, Canhoto, Daniel, & Meadows, 2014).

Hotel accommodation data—provided by licensed tourist accommodation structures (hotels, classified furnished apartments, youth hostels) to government authorities (Ministry of Tourism, Ministry of Interior)—allows us to know information regarding number of nights, occupancy rates, and average prices applied, as well as customer data such as country of origin (based on passport or ID cards) and purpose of visit (work, medical treatment, leisure, etc.). Analyzing this data enables examination of key performance indicators in the hotel sector, similar to what the Sultanate of Oman publishes via its national open data portal regarding tourism activities, including information on licensed tourism institutions distributed by activity type, governorate, and province. The aim is to highlight the size and distribution of tourism institutions and their role in supporting the national tourism sector and enhancing economic development (National Open Data Portal, 2026). This data can be integrated with border crossing data to estimate the proportion of informal tourism (off-the-books activities) by comparing with bookings made through unofficial channels.

Also in this area, we find data related to permits and licenses issued by government authorities (Ministry of Tourism, municipality, Ministry of Environment, etc.) that allow natural or legal persons to practice specific tourism activities (tour guide, hunting trip organizer, diving trip supervisor, etc.) or to operate and manage tourism facilities (travel agencies, hotels, guesthouses, etc.). This data provides insight into the diversity of available tourism services and the readiness of a tourist destination in terms of establishing specific types of tourism activities. For example, the Government of Mauritius provides via its official open data portal data on licenses issued by the Tourism Authority, covering multiple categories such as tourism activity licenses (aquarium, diving, karting, etc.), tourist restaurant licenses, travel agency licenses, and other permits (Opendata mauritius, 2026).

This data also allows estimating carrying capacity—the maximum number of individuals who can use a tourist site without causing unacceptable impacts on environmental resources, based on environmental and physical criteria while still meeting tourist requirements (Sharma, 2016). It also allows measuring current actual usage and comparing it with recommended limits. A scientific study on

Italy's Tremiti Islands—which include a 1,466-hectare marine reserve in the Gargano National Park—revealed a 7% increase in Italian visitors between 2005 and 2021. However, the number of nights spent by tourists decreased by 18%, confirming a shift toward short-stay (day-trip) tourism that places greater pressure on local resources (RanieriI, Santoro, Grilli, De Lucia, & Pazienza, 2024). This may allow analysis and monitoring of violations and discrepancies between permits obtained and actual practices, as well as development of mechanisms to ensure tourism growth aligns with legal and environmental standards (Hadi Aulia & Mulyadi, 2025).

On another front, data generated by electronic financial transactions, credit cards (Visa, Mastercard), central banks, and point-of-sale systems for tourism products (hotels, restaurants, museums, shops, etc.) allows us to know not only the number of arrivals and amounts spent during tourist stays but also the forms and patterns of consumer spending by domestic and international tourists. This spending can be classified according to tourism product basket components (accommodation, food, transport, entertainment, etc.), helping local partners (hotels, restaurants) improve their offerings on one hand and guiding marketing campaigns toward activities most popular among tourists on the other.

For example, the State of Arizona's experience with Visa through the "Visa Destination Insights" platform demonstrates this clearly. The Arizona Office of Tourism moved beyond traditional statistics to achieve detailed analysis of international tourist spending patterns and changing preferences. The data revealed a growing trend toward outdoor activities and natural landmarks rather than urban centers, prompting decision-makers to precisely reallocate marketing budgets toward the most attractive destinations, thereby increasing return on investment (ROI) and improving resource distribution efficiency. This data also supported local businesses by providing accurate visitor profiles, helping them adapt services to meet actual tourist demands. This case represents an ideal model of how financial transaction data can be integrated with border flow data to develop the competitiveness and sustainability of tourist destinations (Visa, s.d.).

Finally, public transport and mobility data tracks tourist movement within destinations using both official and unofficial means of transport. This includes entry and exit points from metro systems or buses (via smart tickets or contactless cards), taxi routes and ride-hailing apps (Uber, Careem), and public bicycle ridership statistics (bike-sharing systems). This data can be obtained from public transport authorities (such as the RATP network in Paris) or from GPS devices installed on buses and taxis. This information helps analyze tourist traffic congestion, identify peak times and locations of greatest pressure, and assists in



designing alternative routes for tourist buses or adding temporary transport stops during festivals and conferences.

The case of the Spanish coastal city of Costa Dorada is worth mentioning. The public transport sector there faces economic challenges due to demand imbalances caused by sharp seasonal tourism shocks. A study analyzed data from 34,641 smart cards with multi-person fares during the summer season of 2018 to analyze public transport user behavior and classify them into profiles based on consumption patterns (temporal activity and spatial distribution). The aim was to develop a differential pricing strategy suitable for each profile and improve bus allocation efficiency, especially during peak tourist times, thereby reducing the financial burden of operating services that exceed actual demand. The study confirmed that smart card data can be used as a predictive model to estimate real-time tourist demand for public transport, helping to dynamically adjust supply (number of buses/operating hours) and thus rationalize operational costs. It provides evidence that Big Data analysis from daily transactions can transform the public transport sector from a financial burden into a service that dynamically adapts to seasonal consumer behavior (Gutiérrez, Domènech, Zaragozaí, & Miravet, 2020).

V. Conclusion:

After reviewing the concept and characteristics of Big Data, analyzing its various sources in the tourism sector, and examining practical applications, we can draw several key findings that confirm the pivotal role of Big Data in transforming tourism management and development.

First: Key Findings Regarding the Nature and Importance of Big Data in Tourism

1. **Multiplicity and Diversity of Sources:** Big Data in tourism comes from official data (border, hotel, licensing), digital platforms (booking sites, review platforms), social media, mobile phone data (CDRs), Internet of Things (sensors, cameras), and financial transactions (credit cards). This diversity requires integrative methodologies (data fusion) to obtain a comprehensive and accurate view.
2. **Transition from Estimation to Precision:** Case studies analyzed (Slovenia's experience using mobile operator data, the Kastoria experience analyzing TripAdvisor reviews, and the Pompeii experience using IoT) prove that Big Data overcomes the limitations of traditional surveys—small sample sizes, high cost, slowness—to achieve real-time, high-precision estimates of tourism flows, behaviors, seasonal patterns, and even tourists' emotional responses.

3. **From Monitoring to Prediction:** The role of Big Data is no longer limited to describing the past or analyzing the present; it has extended to predicting the future. Through machine learning and artificial intelligence models, it is possible to forecast numbers of incoming tourists, demand for specific services, and congestion times, enabling stakeholders to engage in proactive planning and optimize resource allocation.

Second: Most Prominent Practical Applications Revealed by the Study

Based on the experiences presented in this research, modern techniques for handling Big Data allow the development of tourism activities in several key aspects:

1. **Enhancing Tourist Experience:** Big Data contributes to personalizing services (through analyzing individual preferences), improving interaction (through analyzing reviews and comments), and ensuring safety (through smart monitoring and crowd prediction systems).

2. **Tourist Destination Management:** Big Data enables congestion management (via heat maps), improving destination sustainability (by comparing carrying capacity with actual usage), and guiding investments (by analyzing areas with the highest attraction and fewest services).

3. **Tourism Marketing:** Big Data makes it possible to measure the return on investment (ROI) of marketing campaigns, identify the most effective influencers, and precisely target promotional offers to specific segments (based on actual behavior rather than stated intentions).

4. **Economic Decision Support:** Studies such as the Arizona experiment with Visa show how analyzing financial transaction data can redirect marketing budgets toward the most profitable sectors and destinations, and how analyzing public transport data can help rationalize operational costs through dynamic pricing and real-time supply adjustments.

5. **Security and Emergency Management:** The experience of the ancient city of Pompeii demonstrated the ability of IoT and AI-based systems to enhance visitor safety and security through automatic detection of anomalous behavior, continuous monitoring, and rapid emergency response.

Third: Challenges and Future Opportunities

Despite these achievements, exploiting Big Data in tourism still faces major challenges:

1. **Data Quality and Credibility (Veracity):** Errors, duplicate data, and fake reviews require advanced mechanisms for cleaning and verification.

2. **Privacy and Ethics:** The collection and analysis of personal data (especially location and biometric data) raises legal and ethical issues, requiring

compliance with legislation and adoption of transparency and informed consent principles.

3. **Skills Gap:** The lack of qualified human resources to handle Big Data analytics techniques (data scientists, machine learning engineers) is a major obstacle for many tourism institutions, especially small and medium-sized ones.

4. **System Integration:** Integrating data from different sources (governmental, private, platforms) remains a technical and institutional challenge requiring close cooperation between public and private sectors.

Fourth: Research Recommendations

Based on the findings reached, this research recommends the following:

1. Encourage the creation of national open tourism data platforms that aggregate and standardize data from various sources (borders, hotels, transport) and make it available to researchers and decision-makers while respecting privacy conditions.

2. Invest in human capacity building by developing specialized training and educational programs in Big Data analytics and its tourism applications, aimed at professionals and administrators.

3. Establish clear legal and ethical frameworks regulating the collection and use of tourists' personal data, guaranteeing their rights to privacy and information while simultaneously encouraging innovation.

4. Encourage public-private partnerships to finance and develop Big Data projects in tourist destinations, especially those requiring significant investments in technical infrastructure.

Finally, Big Data is no longer a complementary option in managing tourism activity but has become a strategic necessity for achieving leadership and sustainability in an increasingly complex and competitive tourism market. The transition from a "traditional tourism economy" to "smart tourism" will only be achieved through real investment in human and technical capital, while at the same time adhering to the ethical values and human principles that form the essence of the authentic tourism experience.

References:

1. Hadi Aulia, S., & Mulyadi, M. (2025, December 2025). Analysis of Licensing Issues in Tourist Destination Development: Case Study of Hibisc Fantasy Park. (V. G. Arad, Ed.) *Journal of Legal Studies*, 32-45. doi: <https://doi.org/10.2478/jles-2025-0013>
2. Badman, A., & Kosinski, M. (2024, November 18). *What is big data?* Retrieved Decemeeber 2025, 2025, from [www.ibm.com: https://www.ibm.com/think/topics/big-data?mhsrc=ibmsearch_a&mhq=what%20is%20big%20data](https://www.ibm.com/think/topics/big-data?mhsrc=ibmsearch_a&mhq=what%20is%20big%20data)

3. Bruni, A., & Garzia, F. (2025). The IoT/IoE Integrated Security & Safety System of Pompeii Archeological Park. *Applied Sciences*, 15(13). doi:<https://doi.org/10.3390/app15137359>
4. Chen, M., & Strategist, C. (2024, September 23). *What Is Big Data?* Retrieved December 25, 2025, from <https://www.oracle.com/>: <https://www.oracle.com/big-data/what-is-big-data/>
5. DataReportal. (2025, October 15). *Digital 2026 Global Overview Report*. Retrieved mars 22, 2026, from DataReportal: <https://datareportal.com/reports/digital-2026-global-overview-report>
6. Department of Economic and Social Affairs. (2022). *Handbook on Management and Organization of National Statistical Systems, 4th Edition of the Handbook of statistical organisation* (4th Edition ed.). (U. Nations, Ed.)
7. Dibb, S., Ball, K., Canhoto, A., Daniel, E., & Meadows, M. (2014). Taking responsibility for border security: Commercial interests in the face of e-borders. *Tourism Management*, 42, 50-61.
8. Duncan, A., & Hailemariam, A. (2025). Come and say G'day: Using search engine data to understand the dynamics of tourism demand in Australia. *Tourism Economics*, 31(7), 1428-1451. doi:<https://doi.org/10.1177/13548166251352219>
9. ESCWA. (2021). *استخدام البيانات الضخمة في الحساب مؤشرات أهداف التنمية المستدامة في بيروت: مطبوعات الأمم المتحدة المنطقة العربية، التحديات والفرص*.
10. European Union . (2025, 12 06). *Big data*. Retrieved from Shaping Europe's digital future: <https://digital-strategy.ec.europa.eu/en/policies/big-data>
11. FEVAD. (2024, janvier 31). *Le Voyage -Tourisme, premier secteur e-commerce en nombre de clients*. Retrieved from FEVAD: <https://www.fevad.com/le-voyage-tourisme-premier-secteur-e-commerce-en-nombre-de-clients/>
12. Gutiérrez, A., Domènech, A., Zaragoza, B., & Miravet, D. (2020, October). Profiling tourists' use of public transport through smart travel card data. *Journal of Transport Geography*, 88. doi:<https://doi.org/10.1016/j.jtrangeo.2020.102820>
13. ITU. (2021). *اتجاهات التكنولوجيات الناشئة: الذكاء الاصطناعي والبيانات الضخمة لأغراض التنمية. جنيف*.
14. Jenik, C. (2021, Julay 30). *A Minute on the Internet in 2021*. Retrieved mars 20, 2026, from Statista: https://www.statista.com/chart/25443/estimated-amount-of-data-created-on-the-internet-in-one-minute/?srsltid=AfmBOooZ7fXM-HV2UgazqjULn-QDoZOV0Y7jSPzRKm5lOkIklYYu9ABF#google_vignette
15. Lee, G.-C. (2025). A Data-Driven Approach to Tourism Demand Forecasting: Integrating Web Search Data into a SARIMAX Model. (M. AG, Ed.) *DATA*, 10(73). doi:<https://doi.org/10.3390/data10050073>
16. Moghaddam, S. (2024). The Past, Present, and Future of the Internet: A Statistical, Technical, and Functional Comparison of Wired/Wireless Fixed/Mobile Internet. *Electronics*, 13(10). doi:<https://doi.org/10.3390/electronics13101986>



17. Nadiras, L. (2024, fevrier 7). *Le e-commerce dans le secteur du tourisme en France*. Retrieved from Centre national de ressources Tourisme, Académie de Versailles: <https://tourisme.ac-versailles.fr/spip.php?article1881>
18. OCDE. (2015). *Data-Driven Innovation, Big Data for Growth and Well-Being*. Paris: OECD Publishing. Retrieved from <http://dx.doi.org/10.1787/9789264229358-en>
19. OECD. (2025, June 25). *Using mobile network operator data to measure inbound and outbound travel in Slovenia*. Retrieved from OECD: https://www.oecd.org/en/publications/using-alternative-data-sources-and-tools-to-measure-and-monitor-tourism_2a655ab8-en/using-mobile-network-operator-data-to-measure-inbound-and-outbound-travel-in-slovenia_b2aa438e-en.html#section-d1e307
20. Opendata mauritius. (2026, February 2). *Licences issued per Tourist Activities by the tourist authority 2023 & 2024*. Retrieved from Opendata mauritius: <https://data.govmu.org/dataset/licences-issued-per-tourist-activities-by-the-tourist-authority-2023-2024>
21. RanieriI, F., Santoro, D., Grilli, L., De Lucia, C., & Pazienza, P. (2024, janvier 01). *Evaluation of the tourism carrying capacity: the case study of the Tremiti islands*. Retrieved from IRIS - Institutional Research Information System: <https://fair.unifg.it/handle/11369/451873?mode=simple>
22. Sharma, R. (2016). Evaluating total carrying capacity of tourism using impact indicators . (I. S. Association, Ed.) *Global Journal of Environmental Science and Management (GJESM)*, 2(2), 187-196.
23. SILENO. (n.d.). *Big Data*. Retrieved decembre 2, 2025, from SILENO: <https://sileno.com/fr/technologies/big-data/>
24. Visa. (n.d.). *Arizona Office of Tourism attracts international audiences*. Retrieved avril 2026, 2025, from <https://www.visa.com/en-us/business:https://corporate.visa.com/en/customer-stories/visit-arizona.html>
25. Yannacopoulou, A., & Kallinikos, K. (2025). Measuring Destination Image Using AI and Big Data: Kastoria's Image on TripAdvisor. *Societies*, 15(1). doi: <https://doi.org/10.3390/soc15010005>
26. الاتحاد الدولي للاتصالات. (2021). *اتجاهات التكنولوجيات الناشئة: الذكاء الاصطناعي والبيانات الضخمة لأغراض التنمية*. جنيف: ITU.
27. البوابة الوطنية للبيانات المفتوحة, (2026). ديسمبر 13. *بيانات مؤسسات الأنشطة السياحية في سلطنة عُمان*. Retrieved from <https://www.opendata.gov.om/ar/datasets/d9e8fb32-8e88-4bcb-9dce-df263d97a090>.