ASSIGNMENT-1

Big Data Analytics for Sustainable Tourism in the Tourism Industry

A Comprehensive Review and Future Research Directions

for

Big Data Management

Submitted By:

Harsheen Kaur (401906011)
Aditi Puri (402086003)
Lakshit (401908008)
BE-MBA (Class of 2019)

Under the guidance of

Dr. Amit Kumar Bhardawaj
Assistant Professor

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

LMTSM
LM THAPAR SCHOOL OF MANAGEMENT
NEW INDIA STARTS HERE

In partial fulfilment of the requirements for the
Degree of Masters of Business Administration

At

LM Thapar School of Management, Derabassi
Punjab (India)
August 2023

Table of Contents

Big Data Analytics for Sustainable Tourism in the Tourism Industry

A Comprehensive Review and Future Research Directions

Business Domain: Tourism and Hospitality Industry

Market Size

Definition: The Travel & Tourism market contains package holidays, hotel stays, private vacation rentals, camping and cruises. Users represent the number of travellers. Well-known providers of package holidays are online travel agencies (OTAs) such as Expedia and Agoda and tour operators such as TUI. Specialized providers of hotels and private accommodation booked online are, for example, Hotels.com, Booking.com, and Airbnb. The booking volume includes all booked travels made by users from the selected region, independent of the departure and arrival. The scope includes domestic and outbound travel.

Offline bookings made, for example, in a travel agencies office or by telephone are also included in this market.

CAGR (2022-2030) = 5.4%

Market Value (2022) = 11.1 Trillion $

Market Value (2030) = 16.9 Trillion $

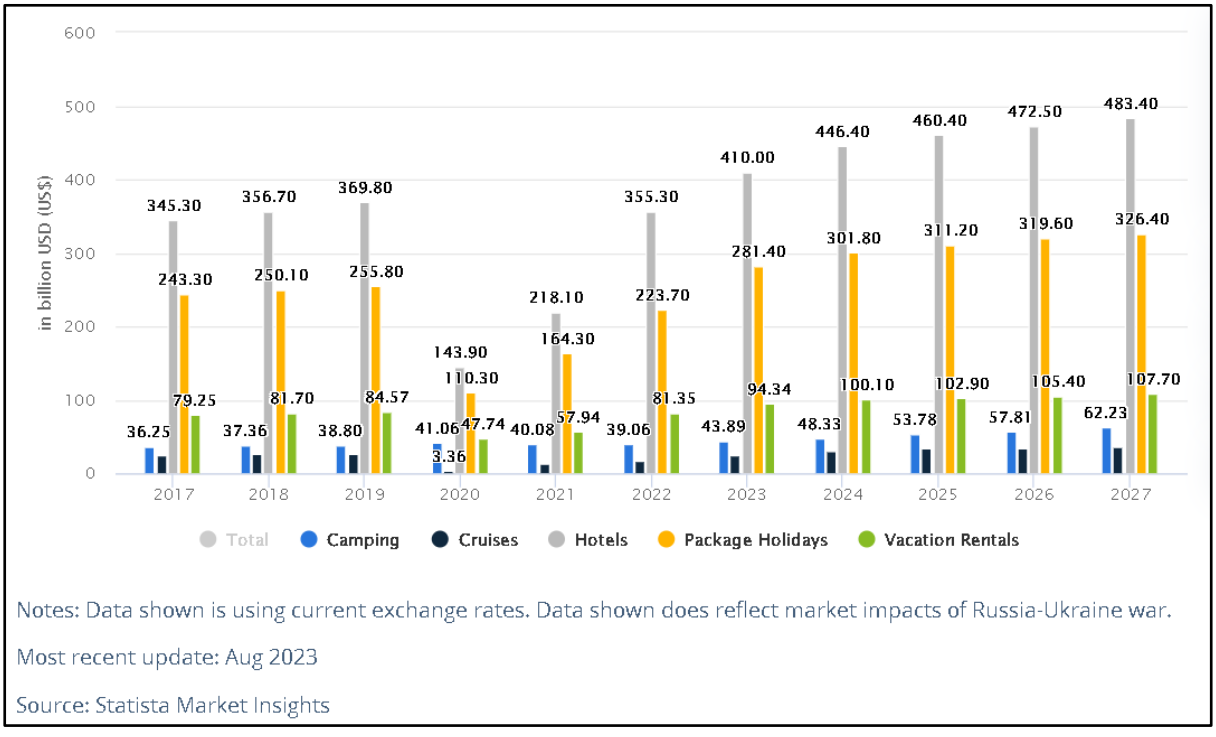


Figure 1 REVENUE OVER YEARS

Impact of Big Data:

- **Customer Personalization:** Businesses can analyse large datasets to understand customer preferences and behaviour. This allows them to offer personalized experiences, from tailored travel recommendations to customized room amenities.
- **Pricing and Revenue Management:** Big data enables businesses to optimize pricing strategies by analysing historical booking data, market demand, and other factors. This helps in maximizing revenue through dynamic pricing models.
- **Operational Efficiency:** Hospitality businesses use big data to improve operational efficiency. For example, hotel chains can use data to manage room occupancy, staff allocation, and energy consumption more effectively.
- **Customer Feedback and Reviews:** Big data analytics can be used to analyse customer reviews and feedback, helping businesses identify areas for improvement and enhance overall customer satisfaction.
- **Marketing and Advertising:** Targeted marketing campaigns can be created using big data to reach potential customers with the right message at the right time. This can lead to better conversion rates and ROI for marketing efforts.
- **Predictive Analytics:** Predictive analytics based on big data can help businesses anticipate future trends and customer behaviour, enabling them to make informed decisions about expansion, marketing strategies, and service improvements.
- **Safety and Security:** In the era of data-driven security, big data is instrumental in enhancing the safety and security of travellers. It can be used for real-time threat detection and emergency response.

Scope

1. **Big data analytics:** Focuses on the application of big data analytics techniques and technologies to address various challenges and opportunities within the tourism industry, particularly related to sustainability.
2. **Sustainable tourism**: Aims to explore and promote sustainable practices within the tourism sector, including environmental, social, and economic sustainability. It seeks to understand how big data can support sustainability efforts in this industry.
3. **Comprehensive review:** The project involves conducting an extensive review of existing literature, research, and case studies related to big data analytics and sustainable tourism in the tourism industry. This review aims to provide a comprehensive overview of the current state of research and practice in this field.
4. **Future research directions:** In addition to the review, the project aims to identify and propose future research directions and potential areas of innovation within the context of big data analytics and sustainable tourism, with a focus on both theoretical and practical applications.

Problem Statement

In the tourism industry, the rapid growth of data generated by various sources, including booking platforms, social media, and IoT (Internet of things) devices, has created both opportunities and challenges for promoting sustainable tourism practices. While there is a

growing interest in harnessing big data analytics to enhance sustainability, there exists a significant gap in the literature in terms of a comprehensive review of the current state of research in this field. This gap hinders the development of a clear understanding of the opportunities and challenges associated with big data analytics in the context of sustainable tourism. Furthermore, there is a need to identify the gaps in existing research and outline future research directions to guide efforts in this critical area. Thus, this study aims to address these issues by conducting a comprehensive review of the literature on big data analytics for sustainable tourism in the tourism industry, with a focus on identifying key research gaps and providing insights into the future research directions required to advance the field.

Project Objectives

The objectives of the project are as follows:

1. To conduct a thorough and systematic review of existing literature and research on the use of big data analytics in the context of sustainable tourism within the tourism industry.
2. To identify the key challenges and opportunities associated with the implementation of big data analytics in sustainable tourism.
3. To examine the current state of big data analytics in sustainable tourism in the tourism industry and to identify the gaps in the existing literature.
4. To propose a framework for big data analytics in sustainable tourism in the tourism industry.
5. To contribute to the overall understanding of the importance of big data analytics in promoting sustainability in the tourism industry.
6. To provide insights into the potential of big data analytics for improving sustainability in the tourism industry, covering aspects such as resource efficiency, environmental impact reduction, and community engagement.
7. To propose a set of future research directions and innovative strategies for harnessing big data analytics to advance sustainable tourism practices and enhance the tourism industry's overall sustainability.

Overall, the project aims to contribute to the knowledge base of big data analytics and sustainability within the tourism industry, providing valuable insights for researchers, policymakers, and industry stakeholders to make informed decisions and promote sustainable practices in tourism.

What are the primary challenges you have identified within the domain of Big Data analytics?

Within the context of our project, "Big Data Analytics for Sustainable Tourism in the Tourism Industry," the primary challenges in the domain of Big Data analytics can include:

1. **Data Quality and Integration**
   Ensuring that the data collected from various sources, including social media, websites, and expert discussions, is of high quality and can be effectively integrated for analysis. Inconsistent or incomplete data can hinder meaningful insights.

2. **Volume and Velocity of Data**
   Dealing with the vast amount of data generated by social media, websites, and expert discussions can be overwhelming. Managing and processing this data in a timely manner is a significant challenge.

3. **Data Privacy and Security:**
   Ensuring the privacy and security of the data collected, particularly when it involves personal discussions or information shared on social media platforms, is crucial to maintain ethical standards and legal compliance.

4. **Complex Data Analysis:**
   Extracting meaningful information from audio and video content (e.g., expert discussions and TED talks) can be challenging due to the complexity of these data formats. Transcription and analysis require advanced techniques.

5. **Sentiment Analysis Accuracy:**
   Achieving high accuracy in sentiment analysis when evaluating the polarity (positive, negative, neutral) of textual data is a challenge, as the context and tone of language can be nuanced.

6. **Bias and Subjectivity:**
   Identifying and mitigating bias in data sources is important to ensure that the analysis is fair and unbiased. Subjectivity in text analysis may also introduce interpretational challenges.

7. **Expert Opinion Variability:**
   The perspectives of experts in the field may vary, and reconciling differing opinions and views in the qualitative analysis can be a complex task, especially when building a conceptual model.

8. **Tools and Software Selection:**
   Choosing the right software tools and platforms for qualitative analysis, as well as ensuring that they are compatible with quantitative data analysis tools, can be a challenge.

Please delineate the primary data sources in context of Man, Machine, and Organization.

1. **Man (People):**

- **LinkedIn Experts and Professionals:** Gather data from LinkedIn experts and professionals in the field of sustainable tourism and Big Data analytics. Conduct interviews, surveys, or collect their opinions and insights to understand their perspectives, opinions, and experiences related to Big Data adoption, sustainability, and the tourism industry.
- **Content from Social Media and Discussion Forums:** Explore social media platforms and discussion forums relevant to the tourism industry. Analyse user-generated content, comments, and discussions to gauge the sentiments, trends, and opinions of people involved in or interested in sustainable tourism and Big Data analytics.
- **Interviews and Expert Discussions:** Collect unstructured data through interviews or expert discussions related to your project topics. This data can provide valuable insights and qualitative information on the challenges, motivations, and perceptions of experts regarding Big Data adoption and sustainability in tourism.

2. **Machine (Technology and Data):**

- **Web Scraping and Crawling:** Utilize web scraping techniques to gather data from websites, blogs, and news articles related to sustainable tourism, Big Data analytics, and the tourism industry. This data can include articles, reports, and textual content that can be analysed.
- **Transcripts of TED Talks and MP3 Discussions:** Extract text from TED Talks and MP3 discussions of experts using automated transcription tools or manual transcription. This will provide you with textual data from these multimedia sources for further analysis.
- **Social Media Data:** Extract data from social media platforms, especially Twitter, to collect tweets and posts related to your project themes. Analyse this data to understand the sentiment and trends in public discussions.

3. **Organization (Businesses and Institutions):**

- **Websites of Tourism Businesses and Organizations:** Collect textual data from the websites of tourism-related businesses, organizations, and governmental bodies. This can include information about sustainability initiatives, tourism statistics, and data-related practices.
- **Academic Papers and Research:** Obtain unstructured data from academic papers and research studies related to the tourism industry, sustainability, and Big Data analytics. Extract text from these sources for analysis.
- **Internal Documents of Tourism Organizations:** If possible, access internal documents of tourism organizations or businesses that discuss their strategies and initiatives related to sustainability and Big Data adoption. Analyse these documents to understand the organizational perspective.

Enumerate six distinct methods for gathering unstructured data.

Project Phase 1: Analysis of Unstructured Data

In the first phase of the project, the objective is to gather and analyse unstructured data from various sources to understand the factors affecting the adoption of big data in the context of sustainable tourism in the tourism industry. The analysis will focus on subjectivity and sentiment analysis using six distinct methods to compare the insights obtained from different platforms. This will help in gaining a comprehensive understanding of the subject matter.

Six Distinct Methods for Gathering Unstructured Data:

1. Image Text Extraction (Google Images):

- Gather 500 images from Google, consisting of 200 related to big data, 200 related to the tourism industry, and 100 related to sustainability in tourism.
- Extract text from these images.
- Create word clouds to visualize word frequency and patterns within the extracted text.
- Analyse the polarity of the text to identify positive, negative, or neutral sentiments.

2. Keyword Analysis (Google and Social Media):

- Collect 2000 relevant keywords from Google and social media platforms like Twitter.
- Create a word cloud to visualize the frequency of these keywords.
- Perform sentiment analysis on the keywords to understand the general sentiment associated with big data, tourism, and sustainability.

3. Audio Discussion Transcription (MP3 Expert Discussions):

- Obtain 10 MP3 discussions from experts on the topics of big data, sustainable tourism, and the tourism industry.
- Use Python to transcribe the audio content into text.
- Generate word clouds from the transcribed text to identify frequently mentioned words.
- Analyse the polarity of the text for sentiment assessment.

4. Video Transcript Analysis (MP4 TED Talks):

- Acquire 10 TED Talks related to big data, sustainable tourism, and the tourism industry.
- Utilize Python to extract and transcribe the spoken content from the videos.
- Develop word clouds to visualize prominent terms in the transcripts.
- Evaluate the sentiment polarity of the text to gauge the opinions expressed in the talks.

5. Web Content Extraction (Websites):

- Identify 10 websites that contain information pertinent to big data, sustainable tourism, and the tourism industry.
- Use Python to scrape and extract text content from these websites.
- Create word clouds to uncover frequent terms and phrases.
- Conduct sentiment analysis to understand the overall sentiment portrayed on these websites.

6. Text Extraction from Articles and Research Papers:

- Select the top 10 articles and research papers related to big data, sustainable tourism, and the tourism industry.
- Extract text from these sources using Python.
- Construct word clouds to highlight key concepts and frequently used terms.
- Analyse the sentiment within the extracted text to discern the opinions expressed in academic literature.

Final Objective:

After employing these six distinct methods to gather unstructured data, the primary aim is to compare the insights obtained from each platform. By analyzing the word clouds and sentiment analysis results, the project aims to understand the subjectivity, sentiment, and key themes prevalent in various data sources. This comparison will provide valuable insights into the factors affecting big data adoption in sustainable tourism within the tourism industry.

Outline your strategy for utilizing unstructured data in qualitative analysis.

To effectively utilize unstructured data in qualitative analysis for our Phase 2, where we intend to derive major themes and develop a questionnaire, we can follow a comprehensive strategy:

Questionnaire Development:

Use the major themes and insights gathered from the unstructured analysis to design a structured questionnaire. Ensure that the questionnaire items are aligned with the themes and aim to collect quantifiable responses related to sustainability, big data, and adoption factors. The questionnaire will include major themes like:

- Opinion about sustainability and role of sustainability in tourism industry
- Opinion about big data and role of big data analytics in tourism industry
- Factors motivating adoption of big data in tourism industry
- Challenges faced while adopting big data in tourism industry

Validation and Iteration:

Pilot test the questionnaire with a smaller group of experts to refine it and ensure that the questions are clear and effectively capture the desired information.

Data Collection:

Administer the questionnaire to a larger sample of LinkedIn experts in the field, and collect structured responses.

Data Preparation and Cleaning:

Gather the responses collected from LinkedIn experts. Ensure the data is organized and free from any irrelevant or redundant information.

Thematic Analysis:

Begin with thematic analysis to identify recurring themes or patterns in the unstructured responses. This will involve coding and categorizing responses into themes related to sustainability, big data, adoption motivators, and challenges.

Interpretive Structural Modeling (ISM) Analysis:

Apply ISM analysis to determine the relationships between different themes and sub-themes. This technique can help in understanding the hierarchy and dependence of these themes.

Fuzzy ISM Analysis:

Use fuzzy ISM to account for the inherent uncertainty in qualitative data. It allows you to incorporate vagueness and fuzziness in the relationships between themes, making the analysis more robust.

Delphi Technique:

Engage experts in a Delphi technique to validate and refine the themes and relationships identified. Delphi involves iterative rounds of feedback from experts, ensuring the credibility of the findings.

Ground Theory Study:

Conduct a ground theory study to develop a comprehensive understanding of the emerging themes. This involves going beyond existing theories to build new, grounded concepts based on the data.

Content Analysis:

Apply content analysis to further examine the content of responses, looking for specific keywords or patterns in the text that may have been missed in the initial thematic analysis.

Narrative Analysis:

Perform narrative analysis to extract and analyse the stories or narratives within the responses. This approach can uncover unique insights and provide a richer understanding of the experts' perspectives.

NVIVO Software:

Utilize NVIVO, a qualitative data analysis software, to assist in organizing, coding, and analyzing the unstructured data. It can help streamline the qualitative analysis process and facilitate collaboration among team members.

Designing the Conceptual Model:

Based on the insights derived from the above techniques, construct a conceptual model that visually represents the relationships between the identified themes. This model should illustrate how sustainability, big data, adoption motivators, and challenges are interrelated within the context of the tourism industry.

Quantitative Analysis:

In Phase 3, use statistical and data analysis techniques to quantify and analyse the structured responses collected through the questionnaire.

By following this strategy, we can effectively transition from unstructured data analysis to the development of a comprehensive conceptual model and questionnaire that will provide valuable insights into the factors affecting big data adoption in sustainable tourism within the tourism industry.

Describe the process of acquiring data from domain experts and explain your intended use of the data.

Sources for data collection:

1. Tweets from X
2. Images and Infographics
3. Expert Interviews
4. Ted talks
5. College Lectures
6. Websites
7. Articles
8. Google Trends

The collected data gathered from above listed sources (text, images, audio, video is firstly to be extracted into text for uniformity in analysis.

Following extraction and transformation of data into text, Word Cloud and Sentiment Analysis are primary processes that will be used to gathered preliminary analysis on gathered data.

Then to be followed using various analysis and studies using models besides hypothesis testing, regression and correlation analysis are to be performed such as:

1. ISM, Fuzzy ISM
2. Delphi method
3. Ground theory study
4. Ethnographic analysis
5. Thematic analysis
6. Narrative analysis

Detail the data collection process for quantitative data. Elaborate on the role of the conceptual model and its key elements.

Some possible types of quantitative data that can be collected:

- Tourist Arrival Data:
  - Number of tourists arriving at specific destinations.
  - Nationality and demographics of tourists.
  - Seasonal and annual variations in tourist arrivals.
- Accommodation Data:
  - Hotel occupancy rates.
  - Average length of stay.
  - Types and categories of accommodations (e.g., budget, luxury).
- Economic Data:
  - Tourism revenue and expenditure.
  - Gross domestic product (GDP) contributions from tourism.
  - Employment statistics in the tourism sector.
- Pricing Data:
  - Average daily rates (ADR) for accommodations.
  - Ticket prices for attractions and activities.
  - Price indices for different tourism-related goods and services.
- Environmental Data:
  - Carbon emissions related to travel and tourism.
  - Water and energy consumption in hotels and resorts.
  - Waste generation and recycling rates.
- Customer Satisfaction and Feedback:
  - Surveys and questionnaires to measure visitor satisfaction.
  - Online reviews and ratings of accommodations, restaurants, and attractions.
  - Net Promoter Score (NPS) data.
- Transportation Data:
  - Passenger numbers for airlines, trains, buses, and cruise ships.
  - Fuel consumption and emissions from transportation modes.
  - Traffic congestion and transportation infrastructure data.
- Cultural and Heritage Data:
  - Number of visitors to museums, cultural sites, and heritage locations.
  - Revenue from cultural and heritage tourism.
  - Historical preservation and conservation data.
- Marketing and Advertising Metrics:
  - Click-through rates (CTR) for online ads.
  - Conversion rates from marketing campaigns.
  - Social media engagement and follower growth.

- Digital Analytics:
  - Website traffic data, including unique visitors and page views.
  - Online booking and reservation data.
  - Conversion rates in e-commerce (e.g., booking conversion rates).
- Employment and Labour Data:
  - Employment data related to the tourism industry, including types of jobs and wages.
  - Labour force participation rates in tourism-related activities.
- Health and Safety Data:
  - Health and safety incidents, such as accidents, illnesses, or injuries among tourists.
  - Data related to the implementation of safety protocols and health measures.
- Supply Chain Data:
  - Data on the supply and demand for goods and services in the tourism industry.
  - Inventory and procurement data for hotels, restaurants, and tour operators.
- Environmental Impact Metrics:
  - Biodiversity and conservation data in ecotourism destinations.
  - Energy efficiency and renewable energy adoption by tourism businesses.
  - Water quality and usage data in coastal and aquatic tourism areas.
- Sustainable Practices Data:
  - Adoption of sustainable tourism practices, such as eco-certifications and eco-labels.
  - Investment in green technologies and initiatives.
  - Sustainability reporting data.
- Competitive Analysis:
  - Market share data for different tourism businesses.
  - Competitive pricing data and strategies.
  - Industry performance metrics.

Define a recommender system and elucidate the benefits you aim to derive from its implementation in your assigned project

A recommender system is a critical component designed to enhance the overall user experience and contribute to the project's objectives. It is a software application or algorithm that analyses large volumes of data to provide personalized suggestions or recommendations to users. These recommendations are typically related to products, services, or content that users are likely to be interested in based on their past behaviour, preferences, and the behaviour of similar users. In the context of the project, the benefits of implementing a recommender system are as follows:

1. **Personalized Travel Recommendations**: A recommender system can analyse a user's historical travel data, preferences, and behaviour to offer personalized travel recommendations. This enhances the user's experience by providing suggestions that are tailored to their individual interests and needs, ultimately leading to more satisfying and memorable travel experiences.

2. **Enhanced User Engagement**: By offering personalized recommendations, the recommender system can increase user engagement with the platform or service. Users are more likely to stay longer, explore more options, and interact with the system when they receive relevant and interesting travel suggestions.

3. **Increased Customer Satisfaction**: Recommender systems can contribute to higher customer satisfaction by helping users find the most suitable travel destinations, accommodations, activities, and experiences. This can lead to positive reviews, repeat business, and word-of-mouth recommendations.

4. **Optimized Resource Allocation**: For the sustainability aspect of the project, a recommender system can assist in optimizing resource allocation. By guiding tourists to less crowded destinations or suggesting eco-friendly options, the system can help reduce the environmental impact of tourism while ensuring that businesses in less-visited areas still thrive.

5. **Revenue Growth**: Recommender systems can lead to increased revenue for businesses within the tourism industry. By suggesting relevant products and services, such as accommodations, tours, or local attractions, businesses can see improved sales and utilization of their offerings.

6. **Data-Driven Decision Making**: The data generated by the recommender system can provide valuable insights into user behaviour and preferences. This data can be used by the project team

and businesses in the tourism industry to make informed decisions about marketing strategies, resource allocation, and sustainability efforts.

7. **Improved Sustainability**: By directing tourists to sustainable and eco-friendly options, the recommender system can contribute to the overall goal of sustainable tourism. It can reduce the environmental impact of travel and promote responsible tourism practices.

8. **Competitive Advantage**: Implementing a recommender system can provide a competitive edge in the tourism industry. Users are more likely to choose platforms and services that offer personalized and relevant recommendations, leading to increased user acquisition and retention.

In summary, a recommender system in the context of "Big Data Analytics for Sustainable Tourism" can play a pivotal role in enhancing the tourism industry's sustainability efforts while simultaneously improving the user experience and driving business success. It achieves this by providing tailored recommendations, increasing user engagement, and promoting responsible and sustainable tourism practices.

Explain how your project's outcomes will be advantageous to academics, the business community, and individuals, including yourself.

The project "Big Data Analytics for Sustainable Tourism in the Tourism Industry" has the potential to generate various advantageous outcomes for academics, the business community, and individuals, including the project team itself. Here's how each group can benefit:

1. Academics:

- **Research Advancement**: The project will contribute to the academic community by providing a comprehensive review of the application of big data analytics in sustainable tourism. This research can serve as a valuable reference for scholars and researchers working in the fields of data science, tourism, and sustainability.

- **Future Research Directions**: The project's exploration of future research directions will offer academics insights into areas where further research is needed. This can guide the development of new research projects and the pursuit of innovative solutions in the intersection of big data and sustainable tourism.

- **Educational Material**: The project's findings and insights can be used as educational material in academic institutions, enriching curricula related to data analytics, sustainability, and the tourism industry.

2. Business Community:

- **Competitive Advantage**: The project's outcomes can provide businesses in the tourism industry with a competitive advantage. By leveraging big data analytics for sustainable tourism, businesses can make data-driven decisions, optimize operations, and attract eco-conscious tourists, leading to increased profitability and market share.

- **Sustainability Practices**: The project's emphasis on sustainable tourism can encourage businesses to adopt environmentally friendly practices. This benefits the industry by reducing its carbon footprint and conserving natural resources, aligning with global sustainability goals.

- **Marketing and Customer Engagement**: Insights gained from the project can help businesses tailor their marketing strategies and customer experiences. They can use data analytics to offer personalized recommendations, enhance customer engagement, and improve customer satisfaction.

3. Tourists:

- **Enhanced Travel Experiences**: Tourists will benefit from more personalized and enjoyable travel experiences. With the project's insights and recommendations, travelers can discover destinations, accommodations, and activities that align with their preferences, resulting in more memorable and satisfying trips.

- **Eco-Friendly Options**: The project's focus on sustainable tourism can guide individuals towards eco-friendly travel choices. This not only contributes to responsible tourism but also allows travellers to connect with nature and local communities in a more meaningful way.

- **Time and Cost Savings**: By providing tailored recommendations, the project can help travellers save time and make informed decisions, potentially saving money and ensuring a smoother travel experience.

4. For my team and I:

- **Professional Growth**: Project team members will gain valuable experience in conducting research, data analysis, and project management. This can enhance their professional skills and open up opportunities for future research or employment.

- **Impact and Recognition**: Successfully implementing this project can lead to recognition within the academic and business communities, boosting the team's reputation and potentially leading to collaboration opportunities or funding for future projects.

- **Personal Satisfaction**: Team members can take pride in contributing to sustainable tourism and making a positive impact on the environment and the tourism industry.

In conclusion, the project's outcomes will have a wide-reaching impact, benefiting academics through research advancements, the business community through competitive advantages and sustainability practices, individuals through enhanced travel experiences, and the project team through professional growth and personal satisfaction. It aligns with the goals of sustainable tourism, economic growth, and responsible data-driven decision-making.

Table of Contents

PROJECT OBJECTIVES

The objectives of the project are as follows:

1. To conduct a thorough and systematic review of existing literature and research on the use of big data analytics in the context of sustainable tourism within the tourism industry.
2. To identify the key challenges and opportunities associated with the implementation of big data analytics in sustainable tourism.
3. To examine the current state of big data analytics in sustainable tourism in the tourism industry and to identify the gaps in the existing literature.
4. To propose a framework for big data analytics in sustainable tourism in the tourism industry.
5. To contribute to the overall understanding of the importance of big data analytics in promoting sustainability in the tourism industry.
6. To provide insights into the potential of big data analytics for improving sustainability in the tourism industry, covering aspects such as resource efficiency, environmental impact reduction, and community engagement.
7. To propose a set of future research directions and innovative strategies for harnessing big data analytics to advance sustainable tourism practices and enhance the tourism industry's overall sustainability.

Overall, the project aims to contribute to the knowledge base of big data analytics and sustainability within the tourism industry, providing valuable insights for researchers, policymakers, and industry stakeholders to make informed decisions and promote sustainable practices in tourism.

CODES USED

R

```
# Import necessary libraries
library(tm)
library(textstem)
library(qdap)
library(dplyr)
library(wordcloud2)
library(tidytext)
library(textdata)
library(SnowballC)
library(wordcloud)
library(RColorBrewer)
library(syuzhet)
library(ggplot2)
library(stringr)

# Data Cleaning
mooncloud <- Corpus(DirSource("directory location"))
inspect(mooncloud)
# Strip unnecessary whitespace
mooncloud <- tm_map(mooncloud, stripWhitespace)
# Convert to lowercase
mooncloud <- tm_map(mooncloud, tolower)
# Remove conjunctions etc.
mooncloud <- tm_map(mooncloud, removeWords, stopwords("english"))
# Remove suffixes to the common 'stem'
mooncloud <- tm_map(mooncloud, stemDocument)
# Remove commas etc.
mooncloud <- tm_map(mooncloud, removePunctuation)
#arguments of 'tm' are converting the document to something other than text, to avoid, run this line
mooncloud <- tm_map(mooncloud, PlainTextDocument)
# Word Cloud
wordcloud(mooncloud
        , scale=c(5,1)     # Set min and max scale
        , max.words=100      # Set top n words
        , random.order=FALSE # Words in decreasing freq
        , shape=round
        , rot.per=0.45       # % of vertical words
        , use.r.layout=FALSE # Use C++ collision detection
        , colors = brewer.pal(8, "Dark2")
)
# Thematic Analaysis
x1 = readLines("C:\\Users\\harsh\\Downloads\\BDM pictures.txt")
# Build a term-document matrix
TextDoc_dtm <- TermDocumentMatrix(x1)
dtm_m <- as.matrix(TextDoc_dtm)
```

```
# Sort by descearing value of frequency
dtm_v <- sort(rowSums(dtm_m),decreasing=TRUE)
dtm_d <- data.frame(word = names(dtm_v),freq=dtm_v)
# Display the top 10 most frequent words
head(dtm_d, 10)
# Plot the most frequent words
barplot(dtm_d[1:10,]$freq, las = 2, names.arg = dtm_d[1:10,]$word,
        col ="cyan", main ="Top 10 most frequent words",
        ylab = "Word frequencies")
# Sentiment Analysis
# please note that different methods may have different scales
syuzhet_vector <- get_sentiment(x1, method="syuzhet")
# see the first row of the vector
head(syuzhet_vector)
# see summary statistics of the vector
summary(syuzhet_vector)
# run nrc sentiment analysis to return data frame with each row classified as one of the following
d<-get_nrc_sentiment(x1)

# head(d,10) - to see top 10 lines of the get_nrc_sentiment dataframe
head (d,10)
#Plot two - count of words associated with each sentiment, expressed as a percentage
barplot(
  sort(colSums(prop.table(d[, 1:10]))), col="yellow",
  horiz = TRUE,
  cex.names = 0.7,
  las = 1,
  main = "Emotions in Text", xlab="Percentage"
)
```

Python

```
import googleapiclient.discovery
import googleapiclient.errors
api_service_name = "youtube"
api_version = "v3"
DEVELOPER_KEY = ""
youtube = googleapiclient.discovery.build(
    api_service_name, api_version, developerKey=DEVELOPER_KEY)

request = youtube.commentThreads().list(
    part="snippet",
    videoId="KENKtzroeK0",
    maxResults=100
)
response = request.execute()
for item in response['items']:
    print(item['snippet']['topLevelComment']['snippet']['textDisplay'])
```

IMAGE DATASET

Methodology:

Data Collection: Collected 250 images related to the topic (big data, big data analytics, sustainability, sustainable tourism, and tourism industry) from Google.

Text Extraction from Images (Python): Used Python libraries like Tesseract or pytesseract to extract text from the collected images. Saved the extracted text into a text file.

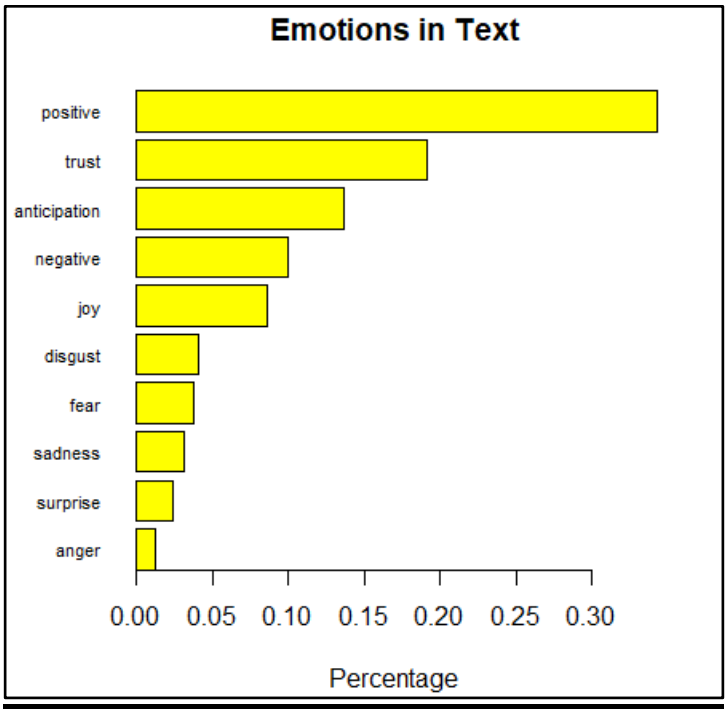
Python code:

```
from PIL import Image
import pytesseract
# Load an image using PIL (Python Imaging Library)
image = Image.open('bdm1.png')
# Use Tesseract to extract text from the image
text = pytesseract.image_to_string(image)
with open('bdm1.txt', 'w') as file:
 file.write(text)
```

Data Preprocessing in R: Loaded the text into R. Performed data cleaning steps such as removing punctuation, special characters, numbers, and stop words. Tokenize the text into individual words or phrases. Lemmatize the text to reduce words to their base forms if needed.

Sentiment Analysis in R: Used sentiment analysis libraries in R (e.g., "tm") to analyze the sentiment of the text. Used pre-trained sentiment lexicons or models. Used to analyse the polarity of the text to identify positive, negative, or neutral sentiments.

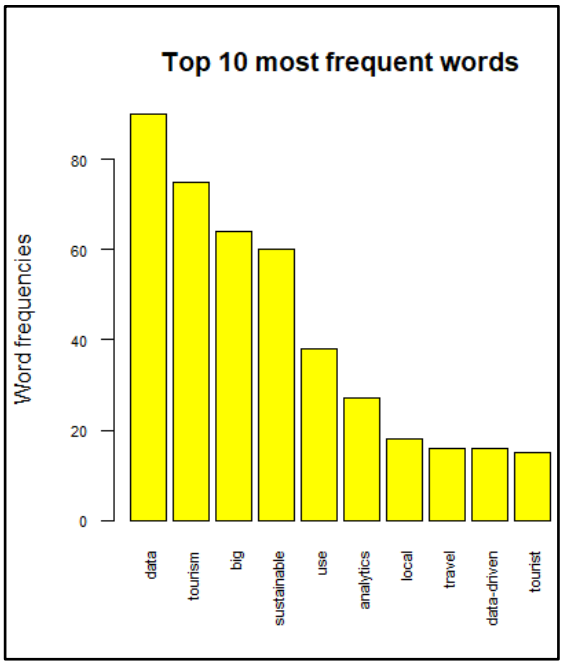
	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
1	0	0	0	0	0	1	0	1	2	0
2	0	1	0	0	1	0	0	1	0	1
3	0	0	0	0	0	1	0	1	1	0
4	0	1	0	0	0	0	0	0	0	2
5	0	1	0	0	1	0	0	1	0	3
6	0	2	0	0	0	1	0	3	1	5
7	0	1	0	0	0	0	0	0	0	4
8	0	2	0	0	1	0	0	1	0	3
9	0	0	0	0	0	0	0	2	0	2
10	1	1	1	1	1	0	0	1	0	2



Word Cloud Analysis in R: Create word clouds to visualize the most common words and phrases in the text file. Word cloud libraries like "wordcloud2" in R can be used for this purpose. Create word clouds to visualize word frequency and patterns within the extracted text.

Thematic Analysis in R: Thematic analysis is a qualitative research method used to identify and interpret themes or patterns within text data.

	word	freq
data	data	90
tourism	tourism	75
big	big	64
sustainable	sustainable	60
use	use	38
analytics	analytics	27
local	local	18
travel	travel	16
data-driven	data-driven	16
tourist	tourist	15



ARTICLES & RESEARCH PAPER DATASET

Methodology:

Data Collection: Selected the top 25 articles and research papers related to big data, sustainable tourism, and the tourism industry. Gathered the abstracts of the top 25 research papers/articles from Google Scholar and Scopus and created a text file containing the collected abstracts.

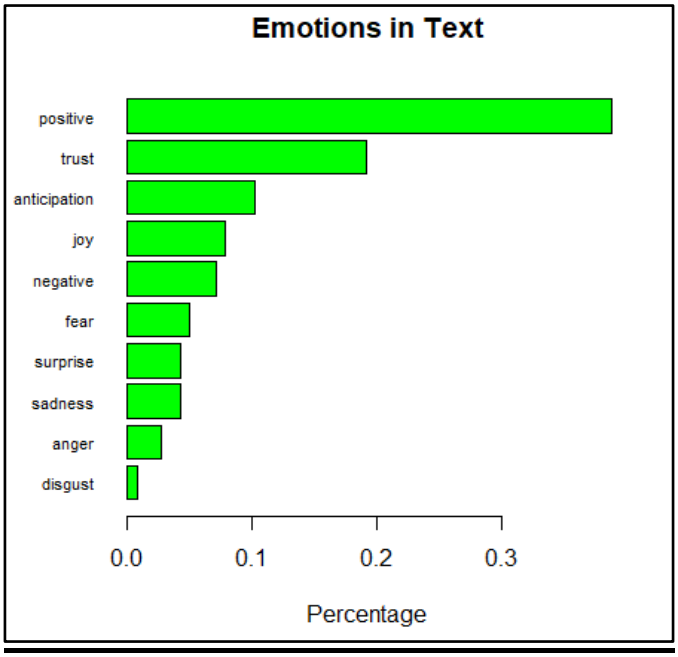
Data Preprocessing: Used R programming language to clean and preprocess the text data. Common preprocessing steps include:

- Removing punctuation, special characters, and numbers.
- Converting text to lowercase.
- Removing stopwords (common words that don't provide much meaning).
- Tokenizing the text (splitting it into individual words or tokens).

Word Cloud Analysis: Created word cloud to visualize the most frequently occurring words in the abstracts. Used R packages like 'wordcloud' or 'tm' for this.

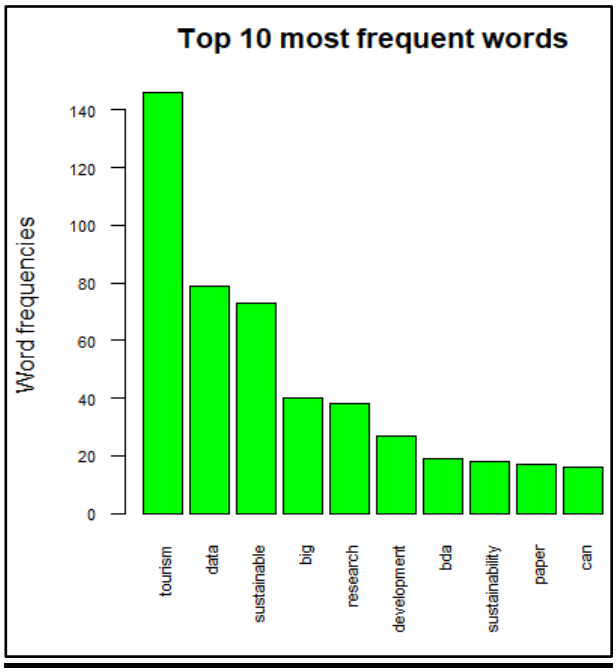
Sentiment Analysis: Utilized a sentiment analysis library or package in R, such as the 'tm' library, to determine the sentiment of each abstract. Sentiment analysis typically categorizes text as positive, negative, or neutral. Calculate the sentiment score for each abstract.

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
1	0	1	0	1	2	0	2	3	2	6
2	2	2	1	0	3	1	1	4	2	8
3	2	1	0	0	0	0	0	3	0	5
4	1	1	0	1	0	0	0	0	2	3
5	2	2	1	2	0	2	0	9	3	5
6	2	3	1	2	3	3	2	7	4	14
7	1	5	0	3	3	3	2	4	3	7
8	1	3	0	2	1	1	2	3	4	7
9	0	3	0	0	4	1	2	4	0	10
10	0	4	0	1	3	1	1	5	0	14



Thematic Analysis: Performed thematic analysis to identify common themes and topics in the abstracts, using R and text mining packages.

	word	freq
tourism	tourism	146
data	data	79
sustainable	sustainable	73
big	big	40
research	research	38
development	development	27
bda	bda	19
sustainability	sustainability	18
paper	paper	17
can	can	16



VIDEO AND AUDIO DATASET

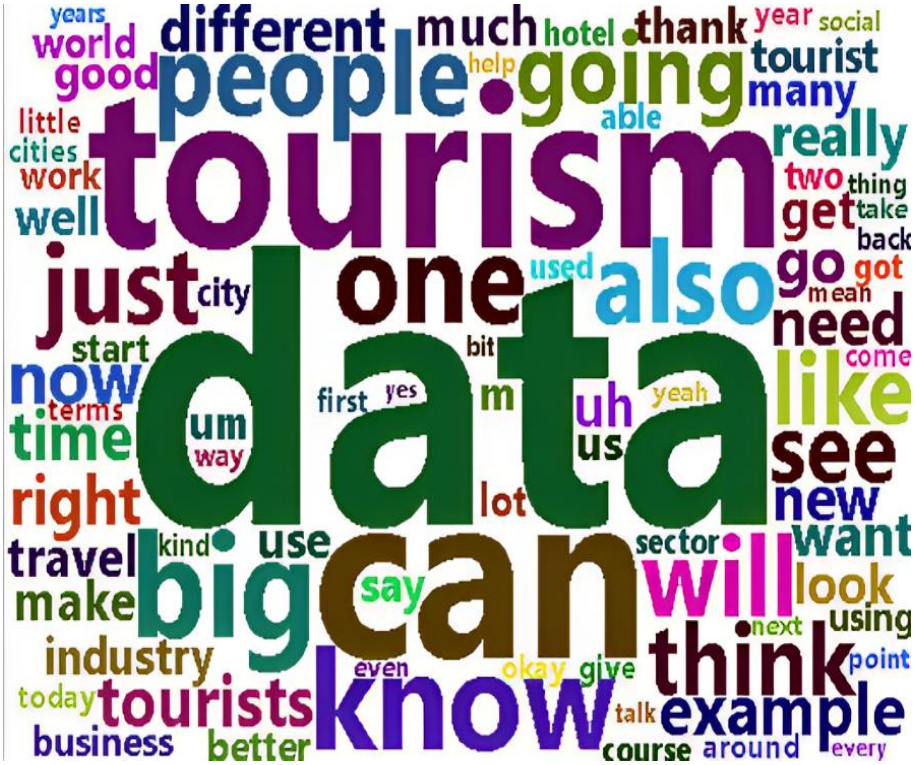
Methodology:

Data Collection: Selected the 25 expert video and audio lectures, discussions and interviews related to big data, sustainable tourism, and the tourism industry.

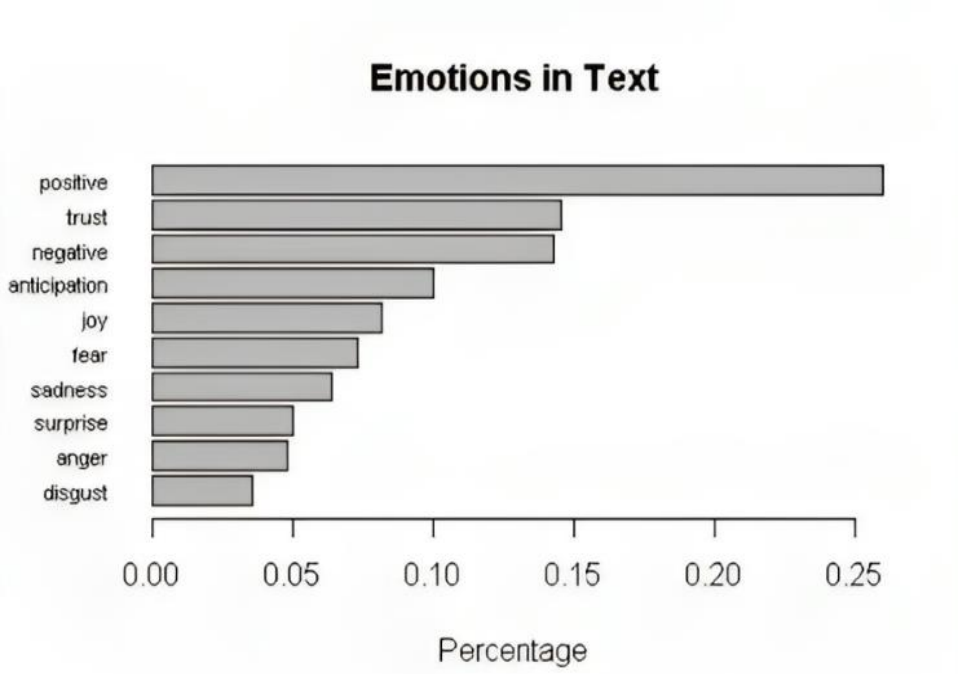
Data Preprocessing: Used R programming language to clean and preprocess the text data. Common preprocessing steps include:

- Removing punctuation, special characters, and numbers.
- Converting text to lowercase.
- Removing stopwords (common words that don't provide much meaning).
- Tokenizing the text (splitting it into individual words or tokens).

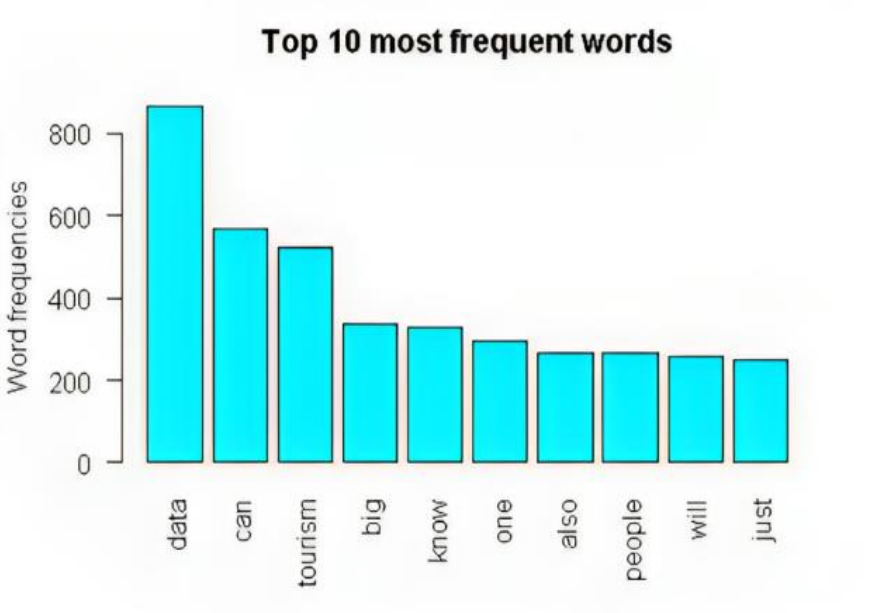
Word Cloud Analysis: Created word cloud to visualize the most frequently occurring words in the abstracts. Used R packages like 'wordcloud' or 'tm' for this.



Sentiment Analysis: Utilized a sentiment analysis library or package in R, such as the 'tm' library, to determine the sentiment of each abstract. Sentiment analysis typically categorizes text as positive, negative, or neutral. Calculate the sentiment score for each abstract.



Thematic Analysis: Performed thematic analysis to identify common themes and topics in the abstracts, using R and text mining packages.



word	freq
the	4145
and	2648
that	1397
you	1257
data	863
this	848
for	780
are	717
can	565
have	543

WEB SCRAPING

Methodology:

Data Collection: Web scraping allows the automated extraction of data from websites, enabling the collection of large volumes of information quickly and efficiently. This data can be used for a wide range of applications, from research and analysis to business intelligence and decision-making.

Businesses can use web scraping to monitor competitors, track pricing information, gather customer reviews, and stay informed about market trends. This data is valuable for making strategic decisions and staying competitive in the market.

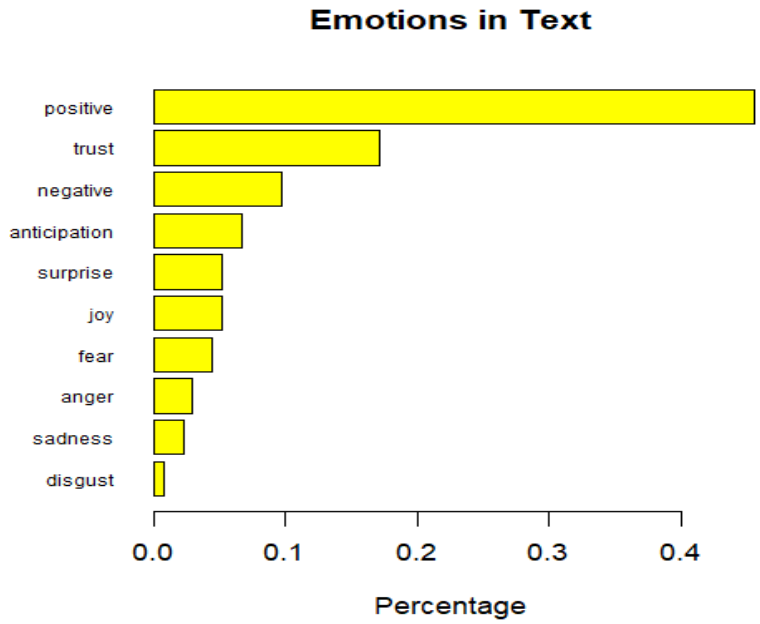
Web scraping provides valuable insights into market dynamics, consumer behavior, and emerging trends. Researchers and marketers can use this data to understand customer sentiment, preferences, and demands.

Data Preprocessing: Used R programming language to clean and preprocess the text data. Common preprocessing steps include:

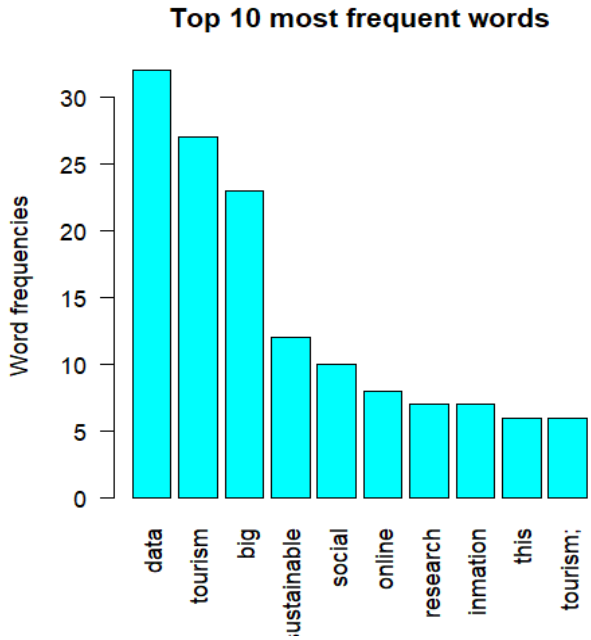
- Removing punctuation, special characters, and numbers.
- Converting text to lowercase.
- Removing stopwords (common words that don't provide much meaning).
- Tokenizing the text (splitting it into individual words or tokens).

Word Cloud Analysis: Created word cloud to visualize the most frequently occurring words in the abstracts. Used R packages like 'wordcloud' or 'tm' for this.

Sentiment Analysis: Utilized a sentiment analysis library or package in R, such as the 'tm' library, to determine the sentiment of each abstract. Sentiment analysis typically categorizes text as positive, negative, or neutral. Calculate the sentiment score for each abstract.



Thematic Analysis: Performed thematic analysis to identify common themes and topics in the abstracts, using R and text mining packages.



TWEETS

Methodology:

Data Collection:

1. Twitter API:

The most common and official way to collect tweets is by using the Twitter API (Application Programming Interface). Twitter provides various API endpoints to access tweet data. You'll need to create a Twitter Developer account and obtain API keys to use this method. Popular libraries for working with the Twitter API in different programming languages include Tweepy (Python) and Twitter4J (Java).

2. Web Scraping:

You can scrape tweets from Twitter's web interface using web scraping tools like BeautifulSoup (Python) or Selenium. Keep in mind that scraping may violate Twitter's terms of service, and it's important to respect Twitter's rate limits and rules when using this method.

3. Third-Party Tools:

Several third-party tools and services are available for collecting tweets, often using the Twitter API under the hood. Examples include Twint, a Python library that scrapes tweets, and TweetDeck, a Twitter-owned tool with advanced search and data export features.

4. Twitter Advanced Search:

You can use Twitter's advanced search functionality to manually search for specific tweets or users. After conducting your search, you can download the results as a CSV file.

5. Streaming API:

The Twitter Streaming API allows you to collect real-time tweets that match specific criteria. This is useful for tracking trends or events as they happen. Tools like Tweepy offer support for the Streaming API.

6. Data Repositories and Marketplaces:

Some organizations and researchers share tweet datasets on data repositories like Kaggle or Zenodo. You can find various datasets related to specific topics or events.

7. Research Collaboration:

Collaborate with researchers who have access to Twitter data or have collected relevant datasets. This may require contacting academic institutions or organizations that have data-sharing agreements with Twitter.

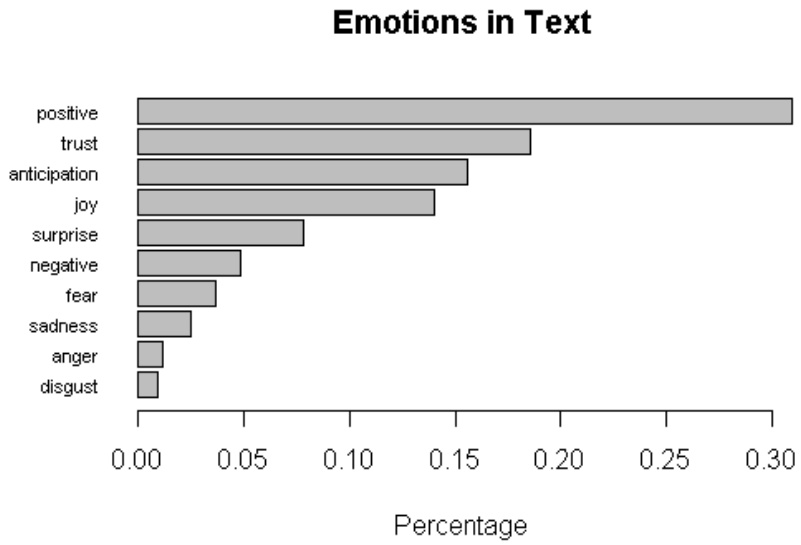
```
import googleapiclient.discovery
import googleapiclient.errors

api_service_name = "youtube"
api_version = "v3"
DEVELOPER_KEY = ""

youtube = googleapiclient.discovery.build(
    api_service_name, api_version, developerKey=DEVELOPER_KEY)

request = youtube.commentThreads().list(
    part="snippet",
    videoId="KENKtzroeK0",
    maxResults=100
)
response = request.execute()

for item in response['items']:
    print(item['snippet']['topLevelComment']['snippet']['textDisplay'])
```

Data Preprocessing: Used R programming language to clean and preprocess the text data. Common preprocessing steps include:

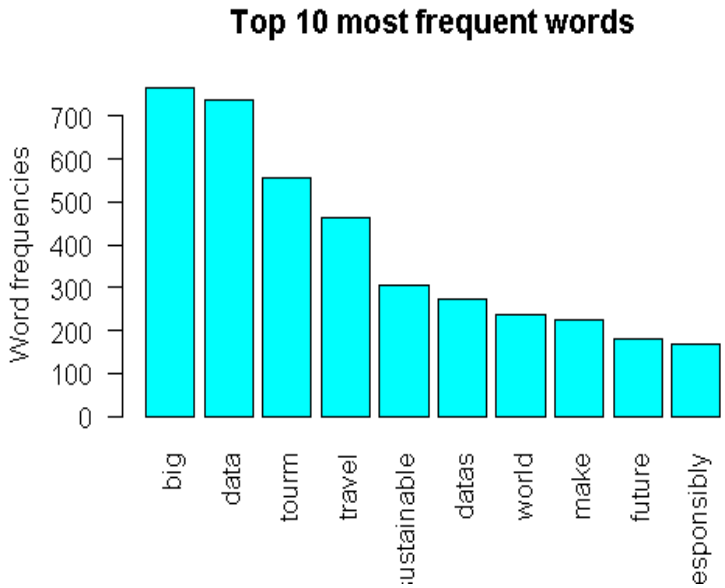
- Removing punctuation, special characters, and numbers.
- Converting text to lowercase.
- Removing stopwords (common words that don't provide much meaning).
- Tokenizing the text (splitting it into individual words or tokens).
-

Word Cloud Analysis: Created word cloud to visualize the most frequently occurring words in the abstracts. Used R packages like 'wordcloud' or 'tm' for this.

Sentiment Analysis: Utilized a sentiment analysis library or package in R, such as the 'tm' library, to determine the sentiment of each abstract. Sentiment analysis typically categorizes text as positive, negative, or neutral. Calculate the sentiment score for each abstract.



Thematic Analysis: Performed thematic analysis to identify common themes and topics in the abstracts, using R and text mining packages.



word	freq
big	765
data	736
tourm	557
travel	462
sustainable	306
datas	272
world	237
make	223
future	182
responsibly	167

COMBINED DATASET

Methodology

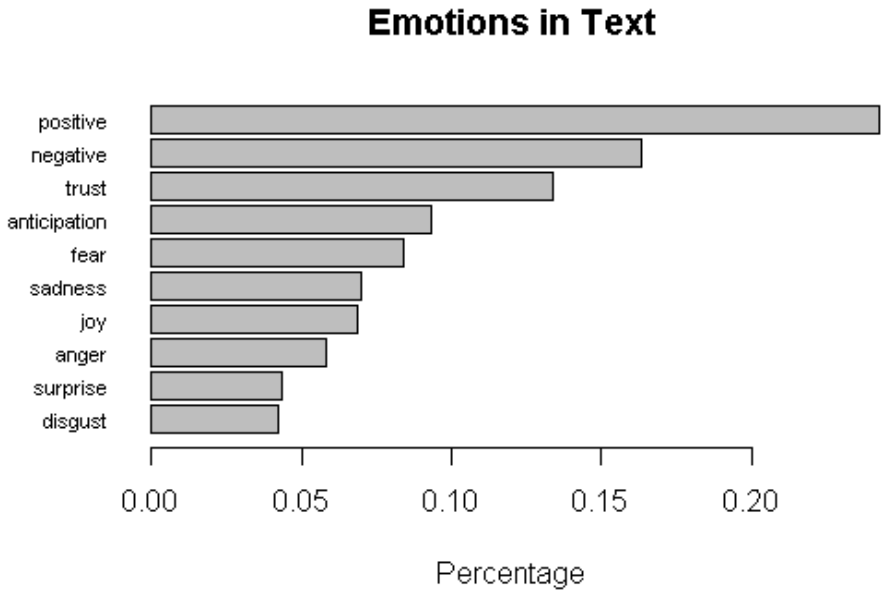
Data Collection: Lastly a single dataset is created by combining all of all the above given datasets, A total of more than 400,000 words is used to create following analysis

Data Preprocessing: Used R programming language to clean and preprocess the text data. Common preprocessing steps include:

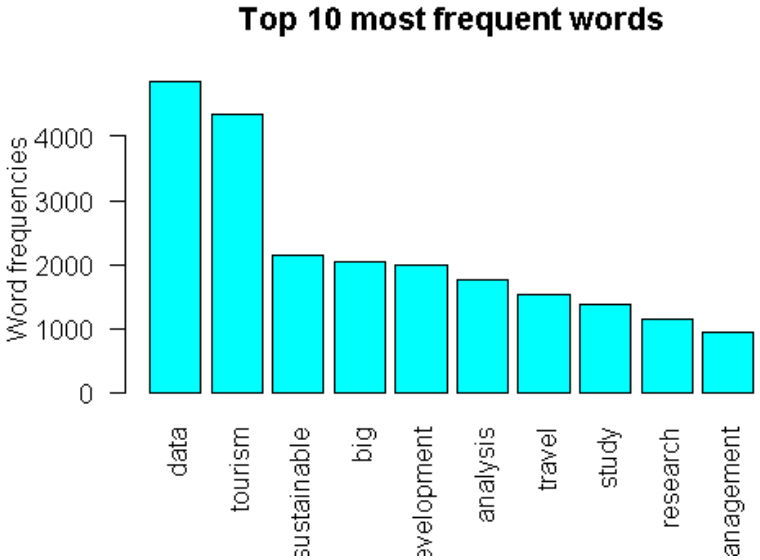
- Removing punctuation, special characters, and numbers.
- Converting text to lowercase.
- Removing stopwords (common words that don't provide much meaning).
- Tokenizing the text (splitting it into individual words or tokens).

Word Cloud Analysis: Created word cloud to visualize the most frequently occurring words in the abstracts. Used R packages like 'wordcloud' or 'tm' for this.

Sentiment Analysis: Utilized a sentiment analysis library or package in R, such as the 'tm' library, to determine the sentiment of each abstract. Sentiment analysis typically categorizes text as positive, negative, or neutral. Calculate the sentiment score for each abstract.



Thematic Analysis: Performed thematic analysis to identify common themes and topics in the abstracts, using R and text mining packages.



word	freq
data	4851
tourism	4334
sustainable	2139
big	2048
development	1996
analysis	1762
travel	1532
study	1367
research	1154
management	937

INTERPRETATION:-

Sentiment Classification:
Sentiment analysis helps classify text into different sentiment categories, providing an understanding of the overall sentiment of a text or a set of texts.

Customer Feedback:
It's often used in customer feedback analysis to understand how customers feel about products or services. For example, analyzing product reviews to determine customer satisfaction.

Brand Monitoring:
Sentiment analysis can be used to monitor social media to understand how a brand is being perceived by the public. This information can inform marketing and public relations strategies.

Trend Analysis:
It's used to track sentiment trends over time, which can be valuable for businesses and researchers to monitor public opinion on various topics.

Automatic Processing:
Sentiment analysis allows for the automatic processing of large volumes of text data, making it efficient for businesses and organizations to understand public sentiment at scale.

Visual Representation:
They provide a visually appealing way to present text data, making it easy to spot the most significant words at a glance.

Data Summarization:
Word clouds are useful for summarizing and presenting text data concisely, making it easier to comprehend and interpret large amounts of text.

Word Frequency:
They indicate the frequency of words, showing which terms are used more frequently in the text.

Article

Big Data Analytics for Sustainable Tourism: A Bibliometric Perspective

Aditi Puri[1]; Harsheen Kaur[2]; Lakshit Adhikari[3]

Abstract: Tourism plays a pivotal role in the global economy, but its sustainability is increasingly threatened by a range of environmental, social, and economic challenges. In the quest to address these challenges and promote sustainable tourism, the integration of Big Data analytics has emerged as a promising approach. This research article presents a comprehensive bibliometric analysis of the field of Big Data analytics for sustainable tourism, shedding light on the current state of research, key trends, and future directions. The study employs bibliometric techniques to analyse a vast array of scholarly publications from various academic databases, providing an overview of the intellectual structure and evolution of the field. It identifies key research themes, influential authors, and prolific journals and institutions contributing to this domain. By examining the co-citation network of academic papers, the research assesses the interconnectivity of research topics, highlighting the most influential concepts and theories in the field. Furthermore, the article delves into the geographical distribution of research, showcasing the global participation and collaboration patterns in the study of Big Data analytics for sustainable tourism. The identification of hotspots and emerging research trends in this field provides valuable insights for policymakers, academics, and industry stakeholders seeking to harness Big Data analytics to advance sustainable tourism practices. The findings of this bibliometric analysis reveal the evolution of Big Data analytics for sustainable tourism from its nascent stages to its current state of maturity, indicating a growing interest in this interdisciplinary domain. The present study uses bibliometric tools and various indicators to discern research progress in the field of tourism industry over the period 2008–2023. Further, VOSviewer software is applied to map the main trends. A total of 1018 publications during the study period were extracted from the SCOPUS database using different keywords related to the big data analytics in tourism industry. This research article contributes to the existing body of knowledge by offering a comprehensive overview of the field's development and providing a roadmap for future research, fostering sustainable tourism practices through the power of Big Data analytics.

Keywords: big data analytics, sustainability, tourism, bibliometric analysis, sustainable tourism practices, literature review

Introduction

In an era characterized by increasing concerns about environmental sustainability and the challenges facing the global tourism industry, the integration of Big Data analytics has emerged as a promising approach to address these pressing issues. Sustainable tourism, with its emphasis on environmental,

[1] Business Analytics and Marketing, Thapar Institute of Engineering and Technology
[2] Business Analytics and Marketing, Thapar Institute of Engineering and Technology
[3] Business Analytics and Operations, Thapar Institute of Engineering and Technology

social, and economic aspects, plays a pivotal role in the global economy. However, it faces a myriad of threats and challenges, ranging from environmental degradation to socio-economic disparities. To navigate these challenges and promote sustainable practices, the field of Big Data analytics has entered the stage, offering innovative solutions and insights. This research article embarks on a comprehensive bibliometric analysis of the intersection between Big Data analytics and sustainable tourism, shedding light on the current state of research, key trends, and future directions.

Tourism, once considered a panacea for economic development, now stands at a crossroads. The rapid growth in tourism activity has placed immense pressure on natural ecosystems, cultural heritage, and local communities. Environmental impacts, such as carbon emissions, habitat destruction, and resource depletion, have raised concerns about the sustainability of tourism. Simultaneously, social issues like overcrowding, cultural commodification, and inequalities have come to the forefront. These challenges necessitate a paradigm shift in the way tourism is managed and understood. Big Data analytics, with its capacity to process vast amounts of data and extract valuable insights, offers a beacon of hope for sustainable tourism. The marriage of Big Data and tourism provides opportunities to address these challenges comprehensively. By harnessing the power of data, decision-makers can gain a better understanding of tourist behaviour, environmental impacts, and socio-economic dynamics. This, in turn, enables the formulation of evidence-based strategies for more sustainable and responsible tourism practices.

Background and Rationale

Tourism is a pivotal global industry, spanning a diverse array of activities and experiences. While it contributes significantly to economic growth and job creation, the sector grapples with numerous sustainability issues. Environmental degradation, socio-cultural disturbances, and economic imbalances have raised critical questions about the long-term viability of tourism. The urgency to safeguard the future of tourism and enhance its sustainability has led to an intriguing fusion of Big Data analytics with this sector. The integration of Big Data offers the potential to revolutionize tourism management, enhance visitor experiences, and contribute to sustainable practices. The use of Big Data analytics within the context of sustainable tourism encompasses a wide spectrum of research and practices. It involves the collection and analysis of vast volumes of data to gain insights into the interactions and impacts of tourism on the environment, communities, and the economy. By harnessing the power of data-driven decision-making, stakeholders in the tourism industry can optimize resource allocation, mitigate negative impacts, and enhance the overall quality of the tourism experience. This research article aims to unravel the intellectual landscape of Big Data analytics for sustainable tourism through bibliometric techniques. We delve into a vast array of scholarly publications sourced from various academic databases, creating a panoramic view of the field's intellectual structure and evolution. We identify key research themes, influential authors, and prolific journals and institutions that are actively contributing to this interdisciplinary domain. The co-citation network of academic papers serves as a compass, guiding us through the intricate interconnectivity of research topics. This network highlights the most influential concepts and theories, shedding light on the intellectual currents shaping the field. Furthermore, this article investigates the geographical distribution of research efforts, showcasing the global participation and collaboration patterns in the study of Big Data analytics for sustainable tourism. The identification of hotspots and emerging research trends provides valuable insights for policymakers, academics, and industry stakeholders eager to leverage Big Data analytics for advancing sustainable tourism practices.

As we traverse the academic landscape of Big Data analytics for sustainable tourism, we witness the evolution of this field from its nascent stages to its current state of maturity. The increasing interest in this domain highlights its relevance and significance in addressing the multifaceted challenges faced by the tourism industry. We utilize bibliometric tools and various indicators to discern the progress of research in the tourism industry from 2008 to 2023. To complement our bibliometric analysis, we employ VOSviewer software to map the main trends, providing a visual representation of the intellectual currents within this field. With a dataset encompassing 1,018 publications from the SCOPUS database over this study period, this research article contributes to the existing body of knowledge by offering a comprehensive overview of the field's development and providing a roadmap for future research.

Scope of the Bibliometric Analysis

This research article aims to explore the dynamic landscape of Big Data analytics for sustainable tourism through a rigorous bibliometric analysis. The study employs sophisticated techniques to analyse a vast corpus of scholarly publications drawn from diverse academic databases. The primary objective is to unravel the intellectual structure and evolution of this interdisciplinary field.

The scope of this bibliometric analysis encompasses:

- Identification of key research themes: Uncovering the dominant topics and themes that researchers have explored in the domain of Big Data analytics for sustainable tourism.
- Influential authors, journals, and institutions: Recognizing the individuals, publications, and academic institutions that have made substantial contributions to this field.
- Co-citation network analysis: Examining the interconnectivity of research topics, thus shedding light on the most influential concepts, theories, and their relationships.
- Geographic distribution of research: Analysing the global participation and collaboration patterns in the study of Big Data analytics for sustainable tourism.

Significance and Implications

The findings of this bibliometric analysis hold tremendous significance for a wide range of stakeholders, including policymakers, academics, industry professionals, and environmental advocates. By offering a comprehensive overview of the development and current state of the field, this research article provides invaluable insights that can shape the future of sustainable tourism practices. The identification of research hotspots, emerging trends, and areas requiring further exploration equips policymakers with the knowledge necessary to make informed decisions regarding tourism regulation and promotion. Academics can use this analysis as a foundation for their research endeavours, leveraging existing knowledge to drive further innovation. For industry professionals, these insights guide strategic planning, helping to harness the potential of Big Data analytics to enhance sustainability in the tourism sector.

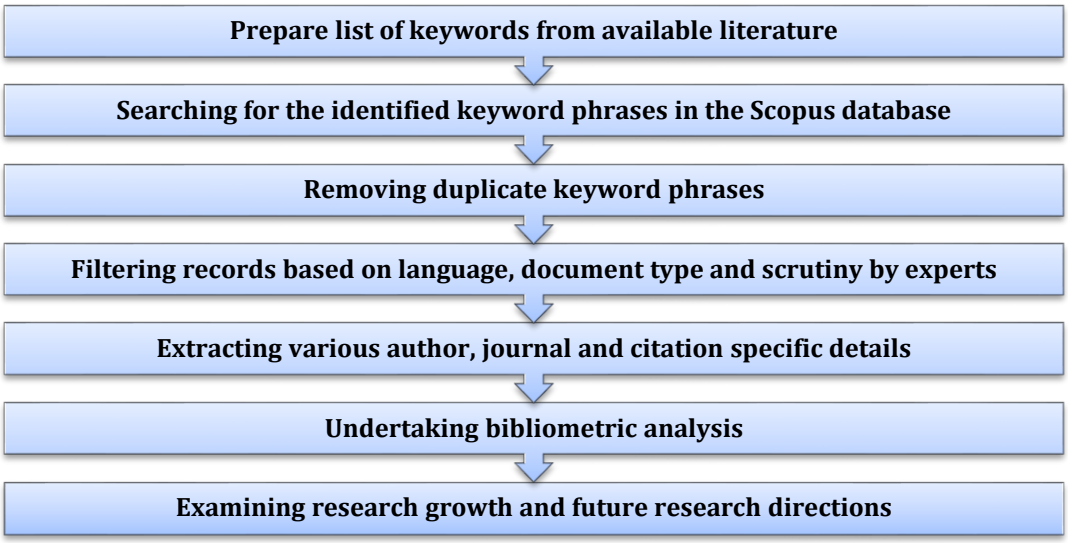
In conclusion, our bibliometric analysis reveals the promising trajectory of Big Data analytics for sustainable tourism. It underscores the growing importance of interdisciplinary approaches and the potential for data-driven decision-making to revolutionize the tourism industry. As we embark on this journey through the academic landscape, we are poised to uncover key insights, trends, and future

directions, ultimately contributing to the advancement of sustainable tourism practices through the power of Big Data analytics. The primary contribution of this research entails a comprehensive bibliometric examination of the body of research on green products and related topics from 1964 to 2019. The study's outcomes encompass an analysis of the growth patterns in publications and citations, identification of the most influential papers, recognition of leading authors, journals, institutions, and countries in this domain, as well as a network analysis of the co-occurrence of keywords supplied by authors pertaining to green products. These findings are expected to provide valuable insights for researchers and readers, facilitating a better understanding of the historical development and current status of the green products field, while also pinpointing potential areas for future research. The subsequent sections of this study are organized as follows: The following section outlines the research methods employed in this study. Following that, the third section presents the results derived from the bibliometric analysis. Subsequently, a discussion of the findings is presented, and the study culminates in the fifth section, which offers a conclusion that includes an exploration of the study's limitations and the direction for future research.

Materials and Methods

To achieve the objectives of this study using bibliometric tools, it is essential to identify, collect, classify, and consolidate the available published knowledge on the chosen topic and related aspects. To do this, we followed an iterative cycle of defining appropriate search keywords, scanning the available resources in the literature, collecting and organizing relevant data, and conducting further analysis using relevant bibliometric tools as suggested by existing studies.

For this study, we employed a similar approach for data collection and evaluation of the literature regarding big data analytics for sustainable tourism over the period of 1964–2019. The collected data was further analyzed using VOSViewer (Visualization of Similarities Viewer, created at Leiden University, Leiden, the Netherlands) software for bibliometric analysis. VOSViewer is one of the most commonly used open-source tools and offers network visualization of authors, institutions, and keywords and their association through cluster analysis [47,48]. This internationally widely used free bibliometric analysis software has been applied by many bibliometric studies [49–51] in the domain of management. For the purpose of this study, a multi-step approach for data collection and analysis was used. This included keyword identification, literature search, removal of duplicate records, data cleansing considering language, document type, and research areas, bibliometric analysis using Microsoft Excel and VOSViewer software, and further analysis to identify the research trends in the field of big data analytics for sustainable tourism.



A list of 100 keyword phrases relevant to the field of big data analytics for sustainable tourism in the domain of Tourism Management was first identified from the literature related to this topic. These keyword phrases were a combination of various adjectives (related to big data and sustainability) and nouns indicating different aspects and attributes (such as tourism, destinations, technology, etc.) in the context of sustainable tourism. An effort was made to include all possible keyword phrases related to big data analytics for sustainable tourism. Further, to ensure the retrieval of all relevant records, various fields like the title of the publications, abstracts of the publications, and author-supplied keywords were searched for all the identified keyword phrases in reputable databases. The search query for the data collection was performed on 27/10/2023. A total of 1027 records were extracted from the chosen database against the keyword phrases.

This approach allowed us to systematically gather and analyse the existing literature in the field of big data analytics for sustainable tourism and offer insights into current research interests and directions for future research.

Big data	195
Sustainable development	188
Tourism	159
sustainability	118
China	92
sustainable development	89
ecotourism	79
tourism development	70
tourism	63
human	60
Sustainable tourism	51
Humans	47
Sustainability	45
Decision making	43
data analysis	42
COVID-19	41

tourist destination	41
Planning	41
big data	35
Smart city	32
article	32
Economics	30
travel behavior	29
tourism management	29
Information management	28
Indonesia	27
spatiotemporal analysis	27
United States	27
Tourism development	26
Data handling	26
data set	26
Surveys	25

sustainable tourism	24
Motor transportation	24
Urban transportation	24
urban transport	24
Smart tourism	23
Data Analytics	23
Tourism industry	23
Traffic congestion	23
urban area	23
Internet of things	22
Article	22
travel	22
adult	22
Artificial intelligence	21
female	21
Public transport	20
Data mining	20
mobility	20
male	20
perception	20
public transport	20
questionnaire survey	20
tourism market	20
Population statistics	19
Ecology	19
Machine learning	18
Sustainable mobility	18
Big Data	18
panel data	18
tourism economics	18
Economic and social effects	18
Travel time	17
social media	17
spatial analysis	17

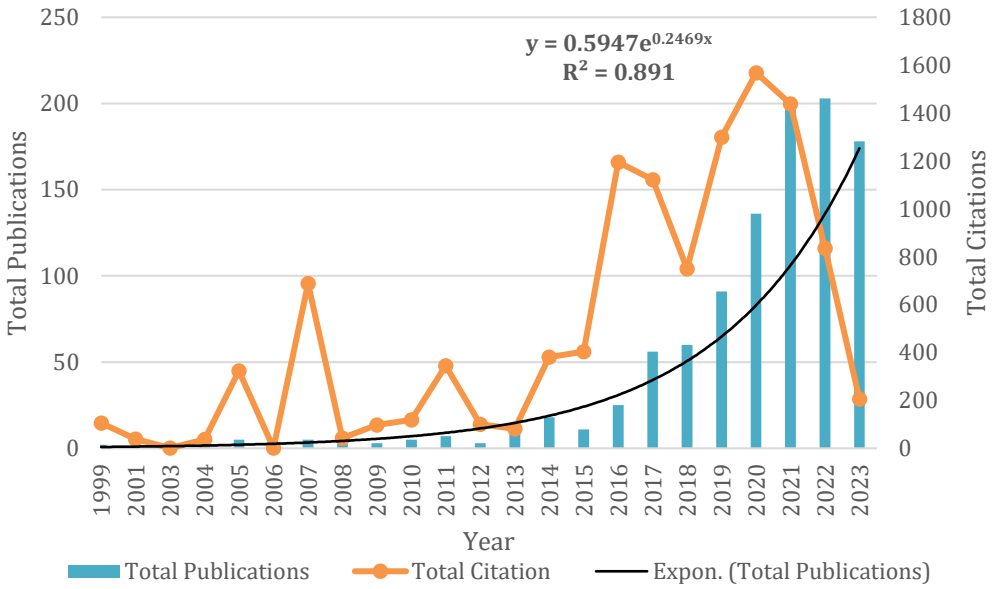
Deep learning	16
Transportation	16
GIS	16
Urban planning	16
Environmental protection	16
Regional planning	16
environmental protection	16
tourist behavior	16
Internet	16
Digital storage	16
Leisure industry	16
machine learning	15
national park	15
economic development	15
Land use	15
Rural areas	15
Male	15
Female	15
Students	15
cultural heritage	14
artificial intelligence	14
Environment	14
Italy	14
Panel data analysis	14
Sustainable Development	14
urban planning	14
marketing	14
Urban mobility	14
South Africa	14
stakeholder	14
Behavioral research	14
regression analysis	14
Information analysis	14
tourist attraction	14

Bibliometric Analysis of Literature: Results

The publications concerning the big data analytics in tourism industry and the related aspects indexed in the SCOPUS database were found distributed in eight document types. The most common format of publication was articles (65.82%), followed distantly by conference papers (21.13%), book chapters (5.93%), reviews (3.89%), and conference review (2.33%). Other forms of publication were less than one percent. It was further found that the records pertaining to green products during the chosen period of 2008-2023 were published in six different languages with English (98.05%) being the most dominant linguistic form of communication followed by Chinese and Spanish.

3.1. Publication Trends of big data analytics in tourism industry Research

The journey of research in the field of big data analytics in tourism industry and related areas is spread over a period of 24 years since 1999. Starting with two publications in the year 1999, it is found that the initial period of 16 years from 1999–2015 has contributed only 7.58% of total publications. However, the last 7 years from 2016–2023 are found to be the most productive as they have contributed to 92.31% of the total publications during the study period of 1999-2023. There has been an exponential growth in the publications related to the area of green products since 2016 (y = $0.5947e^{0.2469x}$ total publications, x = time in years, r^2 = 0.891) as reflected in Figure 2. An exponential curve is fitted on to the data in Figure 2 by using exponential regression. The relative predictive power of the exponential model is denoted by r^2 also known as the coefficient of determination. The value of r^2 (0.891) indicates that 89.1% of the total variation in the number of publications is explained by the relationship between y (total publications) and x (time). Thus, the percentage growth of publications related to big data analytics over time is expected to be 89.10%.



The number of publications is the key to establish the progress of any research field [52]. However, the citations scored by the article since its publication indicate the global impact of the publication [53]. The selected 1027 publications on big data analytics in tourism industry accumulated a total of 11,179 citations with an average citation per paper (ACPP) of 10.89 citations. ACPP is calculated as total citations (TC) divided by total publications (TP). In terms of the year with the highest number of citations, it was found that the year 2020 had received the highest number of citations (1568) for the publications produced in this year with an average citation per paper of 11.52 citations. Out of 25 publications in 2016, nine publications were highly cited with each having 30 or more citations. The citations were taken till October 2023.

3.2. Country Productivity

The analysis of the 1027 records revealed that these publications were published from 69 countries.

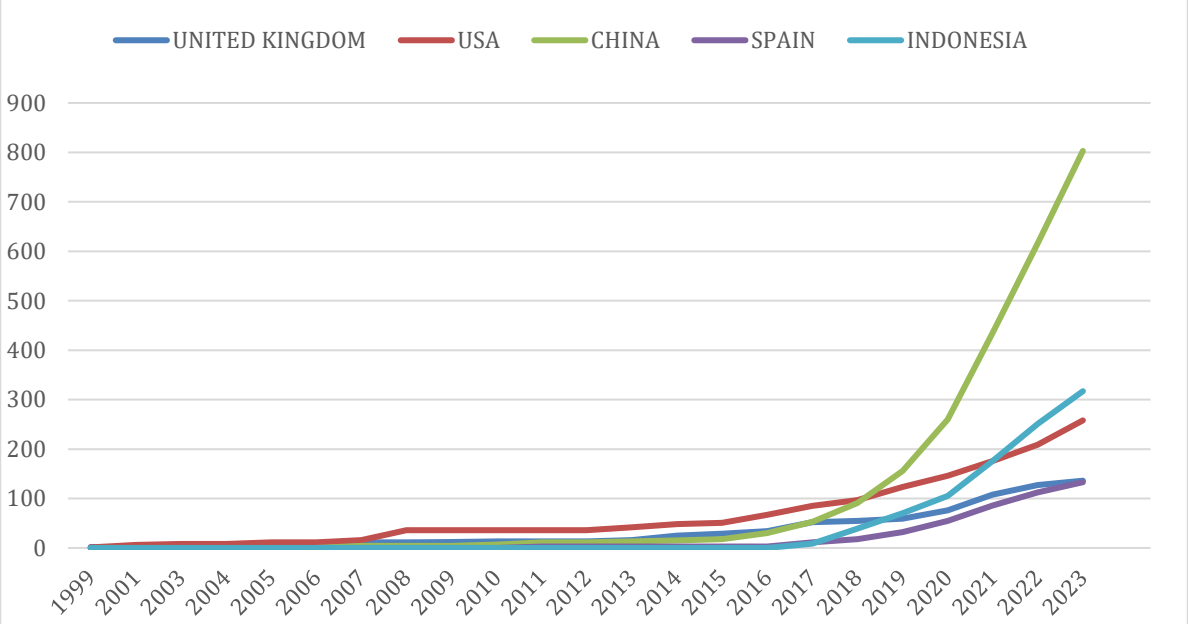
The Table reveals that among the most productive countries producing research related to the big data analytics in tourism industry, the leading country with a contribution of 803 (24.28%) publications out of a total of 3306 was China. China is distantly followed by Indonesia (9.58%), USA (7.80%), United

Kingdom (4.11%), Spain (4.0%), India (3.41%), Portugal (3.02%), Italy (2.90%), Australia (2.66%) and Malaysia (2.57%). As far as total citations are concerned, the top-ranked country USA (2322) was distantly followed by United Kingdom (1105), Spain (754), USA (715), and Australia (651), in that order. United Kingdom (48) was found to occupy the top rank in the case of average citation per paper followed by Singapore (45.7), Belgium (39.2), Canada (34.9), and Pakistan (28.7), in that order. The average citation per paper can be used as a parameter of research valuation, and it is found that the publications from United Kingdom and Singapore are the most frequently cited as compared to those from the most productive countries. This is even though these countries have otherwise a far lower number of publications than the most productive countries.

The top ten most productive countries with 25 or more publications in the field of big data analytics in tourism industry are presented below:

Country	**R(TP)**	**R(TC)**	**R(ACPP)**
CHINA	1(803)	1(2322)	21(10.1)
INDONESIA	2(317)	16(151)	25(2.3)
USA	3(258)	4(715)	13(14.9)
UK	4(136)	2(1105)	1(48)
SPAIN	5(133)	3(754)	10(20.4)
INDIA	6(113)	8(274)	11(17.1)
PORTUGAL	7(100)	11(221)	19(11.1)
ITALY	8(96)	7(307)	14(14)
AUSTRALIA	9(88)	5(651)	8(25)
MALAYSIA	10(85)	10(232)	16(12.9)

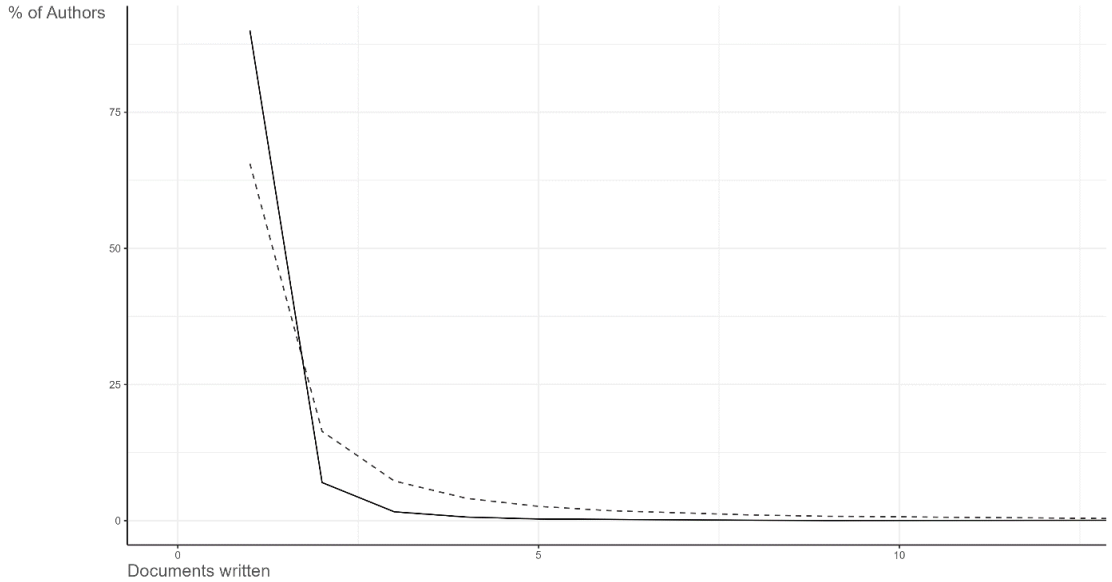
R: Rank; TP: Total Publications; TC: Total Citations; ACPP: Average Citation Per Paper



It is found from Figure 3 that authors from the United Kingdom were publishing in the area of big-data analytics in tourism industry since the year 1999. The first publication from USA also appeared in 1999 and that from the China appeared in 2007. Further, the authors from Spain were publishing since 2011. The publications from the Indonesian authors appeared only in 2017.

3.3. Productive Authors

This section discusses the author's productivity in the area of big data analytics in tourism industry. It is found that 1027 publications were contributed by 3017 authors either singly or in joint authorship. It is also revealed that 24 publications on big data analytics in tourism industry were without author details. Further, out of the remaining 1007 publications, 144 were single-authored, whereas the rest of the 863 publications were multi-authored. The author productivity for single or multi-authored 1007 publications is present in Figure 4.



Authors	Articles	Articles Fractionalized
ZHANG J	13	3.39
ZHANG Y	13	4.17
LIU Y	9	2.45
WANG Y	8	1.93
ZHANG H	8	4.27
CARVACHE-FRANCO M	6	1.30
CARVACHE-FRANCO O	6	1.30
CARVACHE-FRANCO W	6	1.30
CHEN Z	6	1.61
LI J	6	1.58

The table represents the top ten authors who have published the highest number of articles in a particular field, along with their articles fractionalized. Zhang J and Zhang Y ranked first with 13 articles

each, with Zhang Y having the highest articles fractionalized at 4.17. Liu Y ranked third with 9 articles and an articles fractionalized value of 2.45. Wang Y, Zhang H, Carvache-Franco M, Carvache-Franco O, Carvache-Franco W, Chen Z, and Li J all shared the same number of articles (6), but with different fractionalized values ranging from 1.30 to 1.93. This table indicates that Zhang J and Zhang Y have contributed significantly to the field and have been effective in publishing their research, with Zhang Y having the highest fractionalized value, indicating a high level of impact and influence in the field. Liu Y also contributed significantly, ranking third in the list. Furthermore, the table emphasizes the importance of having multiple authors working collaboratively to produce peer-reviewed research with a high impact, as the fractionalized values show that multiple authors have contributed to the same articles.

3.4. Productive Institutions

Affiliation	Articles
ZHEJIANG UNIVERSITY	21
UNIVERSITAS NEGERI MALANG	18
UNIVERSITY OF LJUBLJANA	17
UNIVERSITY OF BRITISH COLUMBIA	16
UNIVERSITY OF JOHANNESBURG	16
UNIVERSITY OF NOVI SAD	16
IPB UNIVERSITY	15
THE HONG KONG POLYTECHNIC UNIVERSITY	15
PEKING UNIVERSITY	14
BEIJING NORMAL UNIVERSITY	13

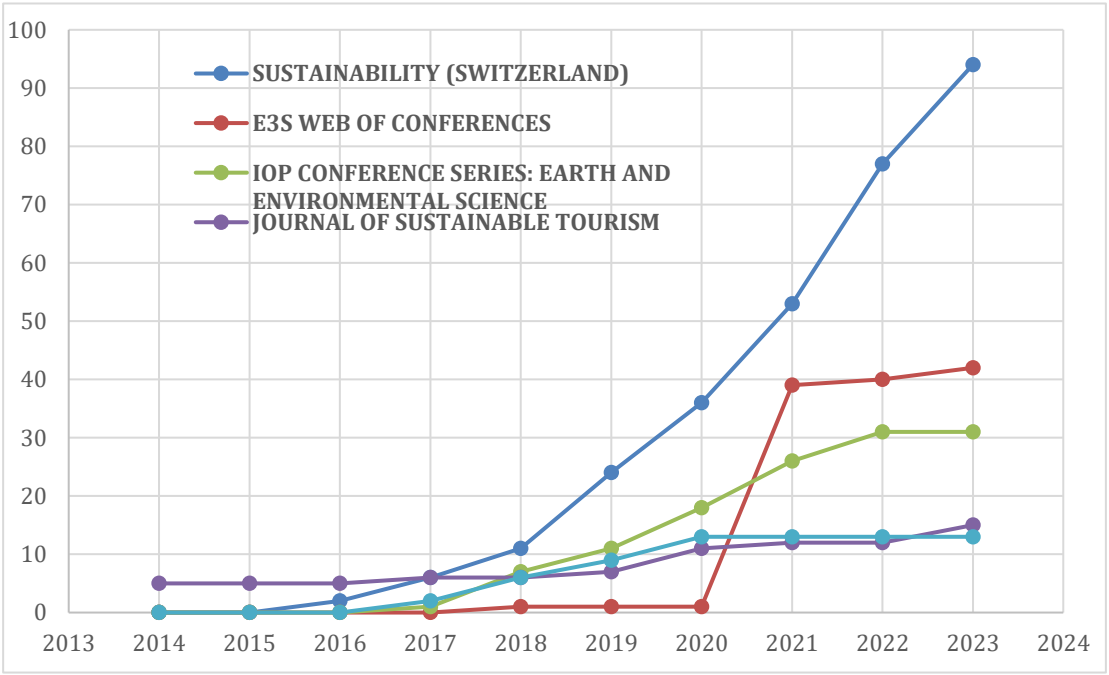
The table shows the top ten universities with the highest number of affiliated articles. Zhejiang University ranked first with 21 articles, followed closely by Universitas Negeri Malang and University of Ljubljana with 18 and 17 articles respectively. The University of British Columbia, University of Johannesburg, and University of Novi Sad all tied for fourth place with 16 articles each. IPB University and The Hong Kong Polytechnic University both had 15 affiliated articles, while Peking University and Beijing Normal University wrapped up the list with 14 and 13 articles respectively. This table indicates that these universities have played a significant role in producing research output, with a significant number of researchers utilizing their resources and collaborating with their faculty members. Furthermore, this highlights the importance of collaboration between institutions in producing high-quality research that can be used to address various issues in society. In conclusion, these top-ranking universities have demonstrated their potential in producing research output and should continue to prioritize initiatives that encourage research collaboration and cultivate strong research environments.

3.5. Productive Journals

Table 4 presents a list of the most productive journals in the field of big data analytics in tourism industry. The top 10 most productive journals have contributed 243 (23.66%) of the total 1027 publications.

Element	h_index	g_index	m_index	TC	NP	PY_start
SUSTAINABILITY (SWITZERLAND)	20	29	2.5	1195	94	2016
JOURNAL OF SUSTAINABLE TOURISM	10	15	0.4	711	15	1999
INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH	7	12	1	174	12	2017
JOURNAL OF CLEANER PRODUCTION	6	6	0.75	489	6	2016
CURRENT ISSUES IN TOURISM	5	11	0.625	175	11	2016
INTERNATIONAL JOURNAL OF CONTEMPORARY HOSPITALITY MANAGEMENT	4	5	1	118	5	2020
IOP CONFERENCE SERIES: EARTH AND ENVIRONMENTAL SCIENCE	4	6	0.571	53	31	2017
JOURNAL OF TRANSPORT GEOGRAPHY	4	5	0.286	183	5	2010
PLOS ONE	4	6	0.8	46	8	2019
TOURISM ECONOMICS	4	4	1	132	4	2020

It can be observed that the SUSTAINABILITY (SWITZERLAND) was the top-ranked journal in the field of big data analytics in tourism industry, and it published in this area for the first time in the year 2016. It has published 94 articles related to green products with 1195 total citations. Further, it is revealed that the JOURNAL OF SUSTAINABLE TOURISM was the second most productive journal and it initiated publishing in the field of green products from the year 1999. This journal published 15 articles concerning green products with 711 total citations. In terms of the number of publications, the INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH further followed by JOURNAL OF CLEANER PRODUCTION (6), CURRENT ISSUES IN TOURISM(11), and INTERNATIONAL JOURNAL OF CONTEMPORARY HOSPITALITY MANAGEMENT (5), in that order.



3.6. Most Cited Articles

The citedness of an article is a quantitative measure based on the number of citations accumulated by the articles since it is published. The most cited articles in the area of big data analytics in tourism industry have been analysed based on total citations from the year of publication (TC), and Total Citations per year and Normalized Total Citations. Table shows the top five most cited articles on big data analytics in tourism industry.

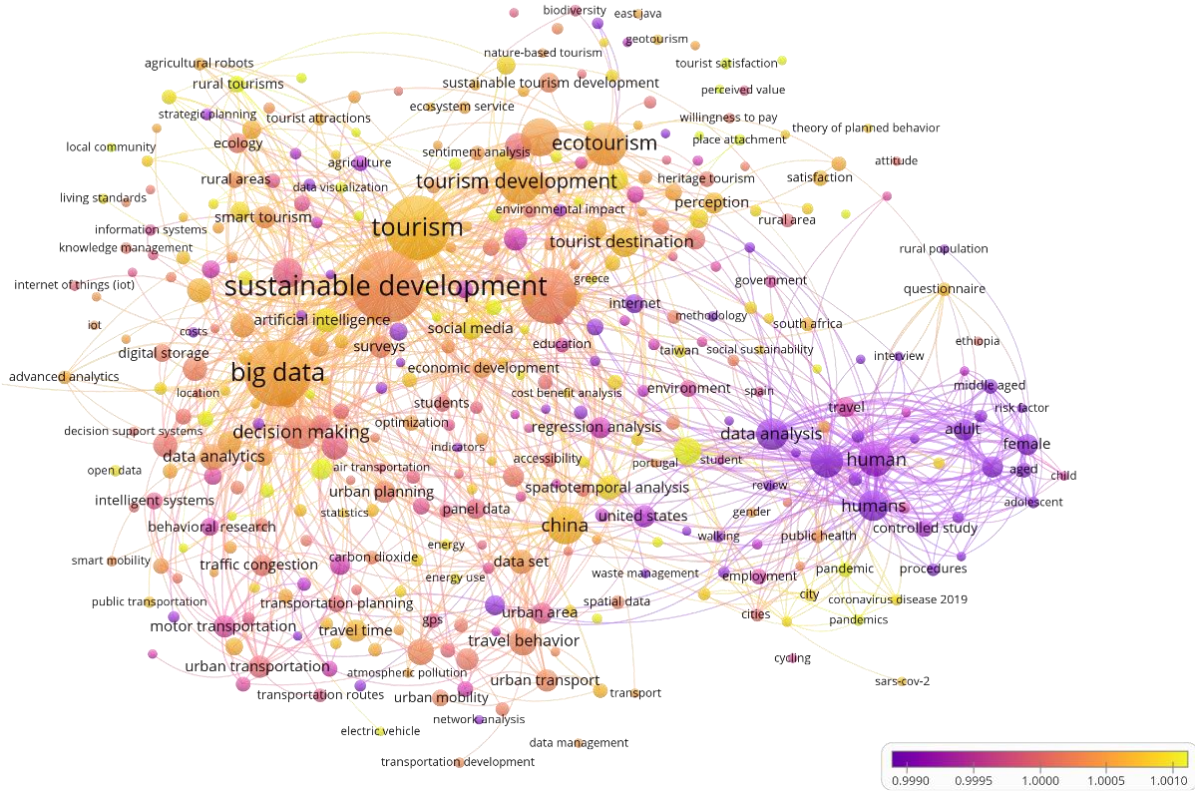
Year	Authors	Title	Cited by	TC per Year	Normalized TC
2016	Sun Y.; Song H.; Jara A.J.; Bie R.	Internet of Things and Big Data Analytics for Smart and Connected Communities	651	81.38	13.62
2007	Ogilvie D.; Foster C.E.; Rothnie H.; Cavill N.; Hamilton V.; Fitzsimons C.F.; Mutrie N.	Interventions to promote walking: Systematic review	518	30.47	3.76
2020	Asadi S.; OmSalameh Pourhashemi S.; Nilashi M.; Abdullah R.; Samad S.; Yadegaridehkordi E.; Aljojo N.; Razali N.S.	Investigating influence of green innovation on sustainability performance: A case on Malaysian hotel industry	199	49.75	17.26
2017	Mathew P.V.; Sreejesh S.	Impact of responsible tourism on destination sustainability and quality of life of community in tourism destinations	183	26.14	9.14
2011	Cowper-Smith A.; de Grosbois D.	The adoption of corporate social responsibility practices in the airline industry	146	11.23	2.96

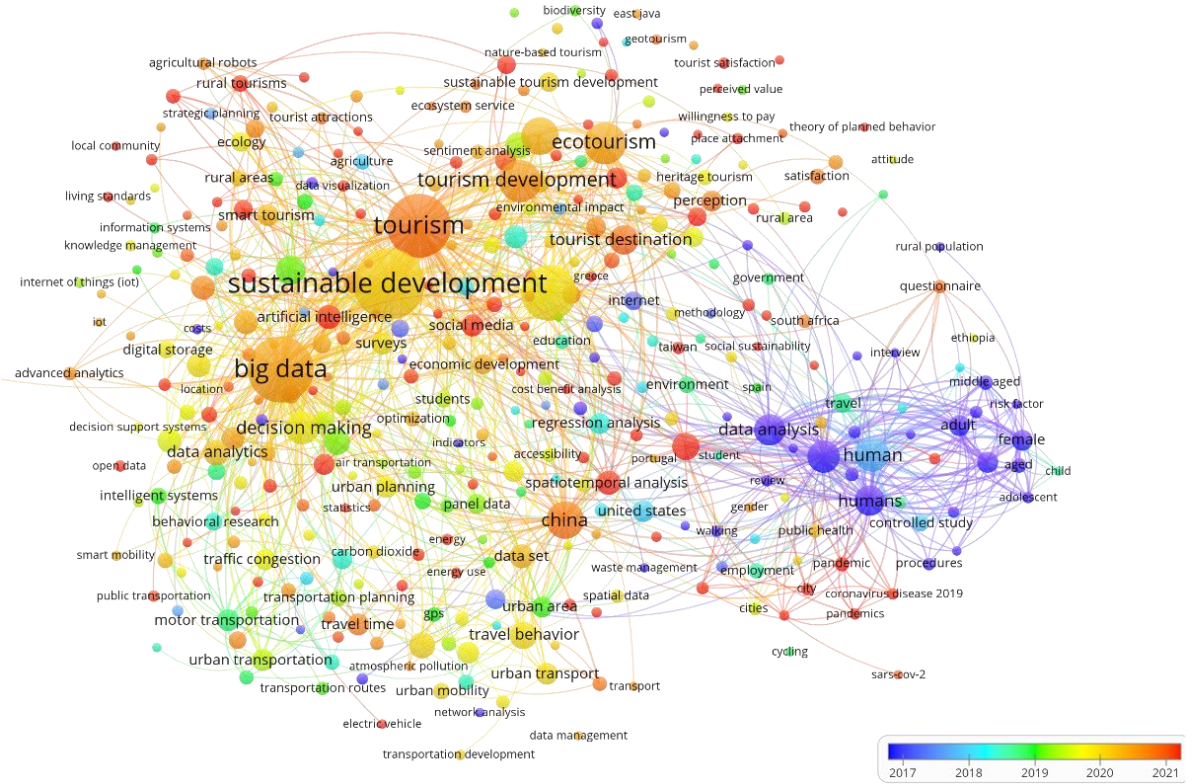
The table presents five research articles published between 2007 and 2020, along with their authors, titles, citation counts, total citation per year (TC per Year), and normalized TC. The first article, published in 2016, titled "Internet of Things and Big Data Analytics for Smart and Connected Communities" by Sun Y. et al., has the highest citation count of 651 and a normalized TC of 13.62. The article explores how IoT and big data analytics can be utilized to develop smart and connected

communities. The second article, published in 2007, by Ogilvie D. et al., titled "Interventions to promote walking: Systematic review," has a citation count of 518 and a normalized TC of 3.76. It reviews different interventions to promote walking in various contexts. The third article, published in 2020, by Asadi S. et al., titled "Investigating influence of green innovation on sustainability performance: A case on Malaysian hotel industry," has a citation count of 199 and a normalized TC of 17.26. It explores the impact of green innovation practices on the sustainability performance of hotels in Malaysia. The fourth article, published in 2017, by Mathew P.V. and Sreejesh S., titled "Impact of responsible tourism on destination sustainability and quality of life of community in tourism destinations," has a citation count of 183 and a normalized TC of 9.14. It investigates the effect of responsible tourism on the sustainability and quality of life of communities in tourist destinations. The fifth article, published in 2011, by Cowper-Smith A. and de Grosbois D., titled "The adoption of corporate social responsibility practices in the airline industry," has a citation count of 146 and a normalized TC of 2.96. It examines the adoption of corporate social responsibility practices in the airline industry.

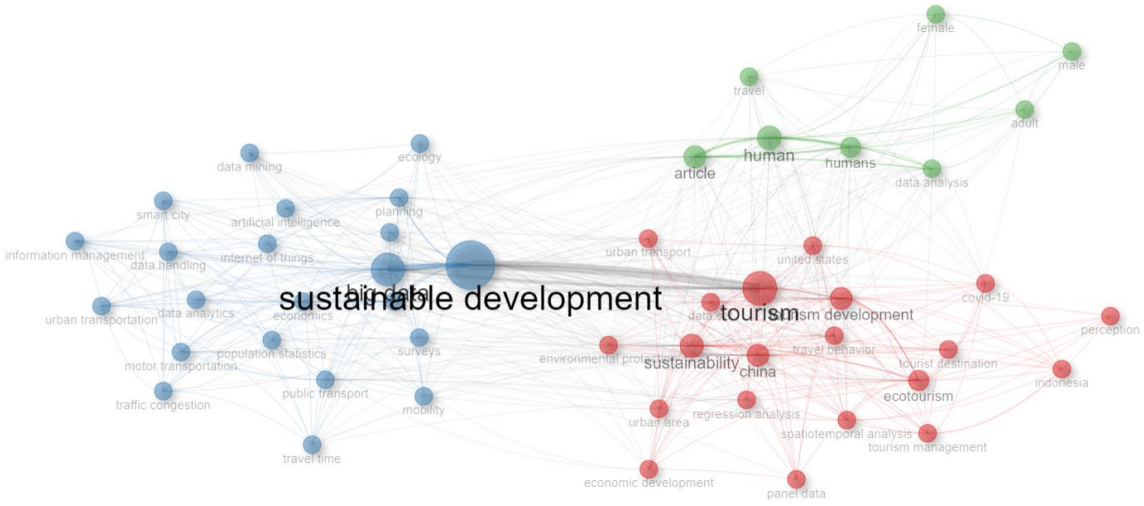
3.7. Author Keywords Analysis

To understand the growth in an area of study, the concept of keyword extraction can be used. The extant literature shows that many bibliometric studies have employed keyword extraction to examine the growth of a subject area. The network diagram is prepared by using VOSViewer software to create a visualization of the co-occurrence of the keyword terms in the domain of the subject area. The co-occurrence is computed as the number of times two keywords appear together in publications. The keyword terms grouped into the same cluster are more likely to reflect identical topics. The keyword most in common is the largest node for that cluster. Further, the changes in the colours of the cluster as one moves from one cluster to another reveal how the area of the study has progressed over the years in the present study, the co-occurrence of keywords in author-supplied keywords has been examined. The minimum occurrence of the words plotted for the mean network diagram is five.


biodiversity
east java
geotourism
nature-based tourism
tourist satisfaction
agricultural robots
rural tourisms
sustainable tourism development
perceived value
ecosystem service
strategic planning
tourist attractions
willingness to pay
theory of planned behavior
local community
ecology
ecotourism
place attachment
sentiment analysis
attitude
agriculture
rural areas
data visualization
tourism development
heritage tourism
satisfaction
living standards
environmental impact
perception
smart tourism
tourism
rural area
information systems
tourist destination
knowledge management
rural population
internet of things (iot)
sustainable development
greece
government
questionnaire
internet
artificial intelligence
methodology
south africa
iot
social media
costs
surveys
education
taiwan
social sustainability
ethiopia
digital storage
big data
economic development
interview
advanced analytics
location
cost benefit analysis
environment
spain
middle aged
students
travel
risk factor
decision support systems
optimization
decision making
regression analysis
data analysis
adult
female
indicators
data analytics
accessibility
portugal
student
human
aged
open data
air transportation
urban planning
spatiotemporal analysis
review
child
intelligent systems
statistics
panel data
humans
adolescent
china
united states
gender
behavioral research
controlled study
energy
walking
public health
smart mobility
carbon dioxide
data set
traffic congestion
waste management
employment
pandemic
procedures
energy use
city
public transportation
transportation planning
spatial data
coronavirus disease 2019
urban area
cities
motor transportation
gps
travel time
pandemics
travel behavior
cycling
urban transportation
atmospheric pollution
urban transport
sars-cov-2
transportation routes
urban mobility
transport
network analysis
electric vehicle
data management
transportation development
0.9990
0.9995
1.0000
1.0005
1.0010

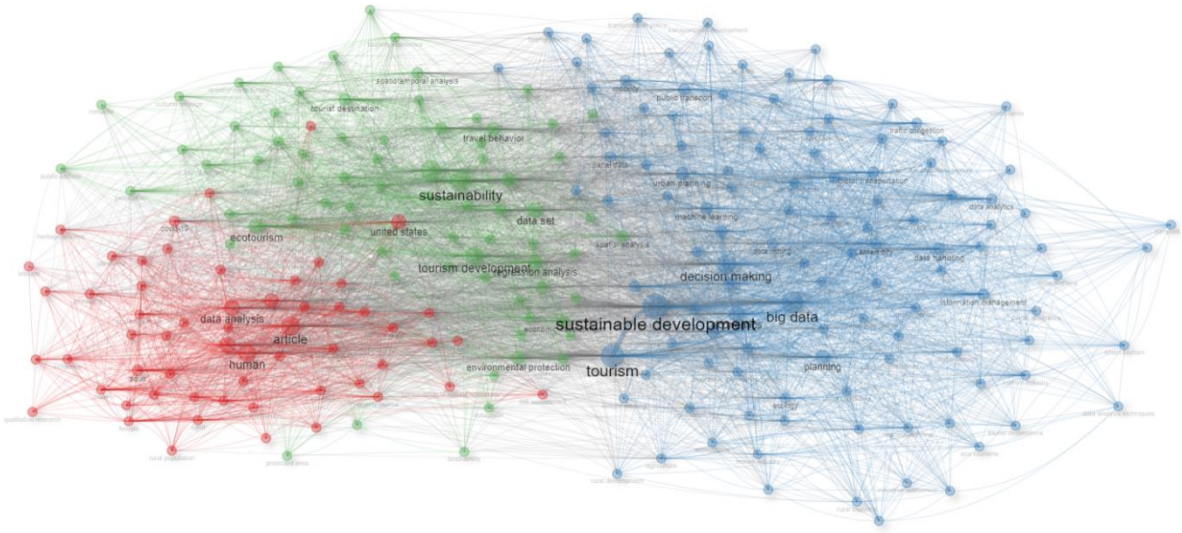

biodiversity
east java
geotourism
nature-based tourism
tourist satisfaction
agricultural robots
rural tourisms
sustainable tourism development
perceived value
ecosystem service
strategic planning
tourist attractions
willingness to pay
theory of planned behavior
local community
ecology
ecotourism
place attachment
sentiment analysis
attitude
agriculture
rural areas
data visualization
tourism development
heritage tourism
satisfaction
living standards
environmental impact
perception
smart tourism
tourism
rural area
information systems
tourist destination
knowledge management
rural population
internet of things (iot)
sustainable development
greece
government
questionnaire
internet
artificial intelligence
methodology
south africa
iot
social media
costs
surveys
education
taiwan
social sustainability
ethiopia
digital storage
big data
economic development
interview
advanced analytics
location
cost benefit analysis
environment
spain
middle aged
students
travel
risk factor
decision support systems
optimization
decision making
regression analysis
data analysis
adult
female
indicators
data analytics
accessibility
portugal
student
human
aged
open data
air transportation
urban planning
spatiotemporal analysis
review
child
intelligent systems
statistics
panel data
humans
adolescent
china
united states
gender
behavioral research
controlled study
energy
walking
public health
smart mobility
carbon dioxide
data set
traffic congestion
waste management
employment
pandemic
procedures
energy use
city
public transportation
transportation planning
spatial data
coronavirus disease 2019
urban area
cities
motor transportation
gps
travel time
pandemics
travel behavior
cycling
urban transportation
atmospheric pollution
urban transport
sars-cov-2
transportation routes
urban mobility
transport
network analysis
electric vehicle
data management
transportation development
2017
2018
2019
2020
2021

3.8. Co-occurrence network



A co-occurrence network is a type of visualization that shows the relationships between different words or phrases in a body of text. In this case, the co-occurrence network shows the relationships between different keywords in the literature on big data analytics and sustainable tourism. The size of the nodes in the network represents the frequency of occurrence of the keywords, and the thickness of the edges between the nodes represents the strength of the relationship between the keywords. The network is colored to represent different clusters of related keywords. The co-occurrence network can be used to identify the key concepts and themes in the literature. For example, the network shows that the term "sustainability" is central to the research, and that it is closely linked to other important concepts such as "data mining", "information management", "data handling", "internet of things", "sustainable development", "human", "article", "humans", "urban transport", "tourism development", "perception", "road transport", "motor transportation", "traffic congestion", "public transport", "sustainability china", "tourist destination", "ecotourism", "regression analysis", "spatiotemporal analysis", "tourism management", and "economic development". This suggests that big data analytics is being used to address a wide range of sustainability challenges in the tourism industry. For example, researchers are using big data to track tourist flows, identify and manage over-tourism, and develop more sustainable tourism practices. The co-occurrence network also reveals some interesting relationships between different research topics. For example, the terms "data mining" and "information management" are closely linked, which suggests that there is a strong focus on developing new data mining techniques for sustainable tourism research. Overall, the co-occurrence network provides a valuable overview of the research landscape on big data analytics and sustainable tourism. It highlights the key concepts and themes in the literature, and it suggests some promising directions for future research.

3.9. Thematic map



Big data analytics has emerged as a powerful tool for sustainable tourism research. By analyzing large and diverse datasets, researchers can gain insights into tourist behaviour, destination management, and the environmental and social impacts of tourism. Bibliometric analysis is a valuable technique for identifying and analyzing the trends and patterns in big data analytics research on sustainable tourism. The thematic map you have provided is a visual representation of the relationships between different keywords and concepts in the literature on big data analytics and sustainable tourism. The map shows that the term "sustainability" is central to the research, and that it is closely linked to other important concepts such as "data set", "united states", "decision making", "article", "sustainable development", and "human". This suggests that big data analytics is being used to address a wide range of sustainability challenges in the tourism industry. For example, researchers are using big data to track tourist flows, identify and manage over-tourism, and develop more sustainable tourism practices.

The thematic map also reveals some interesting relationships between different research topics. For example, the terms "decision malang" and "article" are closely linked, which suggests that there is a growing body of research on how to use big data analytics to inform decision-making in the tourism industry. Overall, the thematic map provides a valuable overview of the research landscape on big data analytics and sustainable tourism. It highlights the key concepts and themes in the literature, and it suggests some promising directions for future research.

Here are some specific theories that are relevant to the thematic map:

- Sustainability theory provides a framework for understanding the environmental, social, and economic impacts of tourism.
- Big data theory provides a framework for understanding the collection, storage, analysis, and visualization of large datasets.
- Bibliometric theory provides a framework for analyzing the patterns and trends in scientific literature.

3.10. Word cloud

The word cloud visually encapsulates the key themes explored in your research paper titled "Big Data Analytics for Sustainable Tourism: A Bibliometric Perspective." This word cloud is an amalgamation of terms that revolve around the realms of sustainable development, tourism, and big data analytics.

The word cloud predominantly features the following terms:

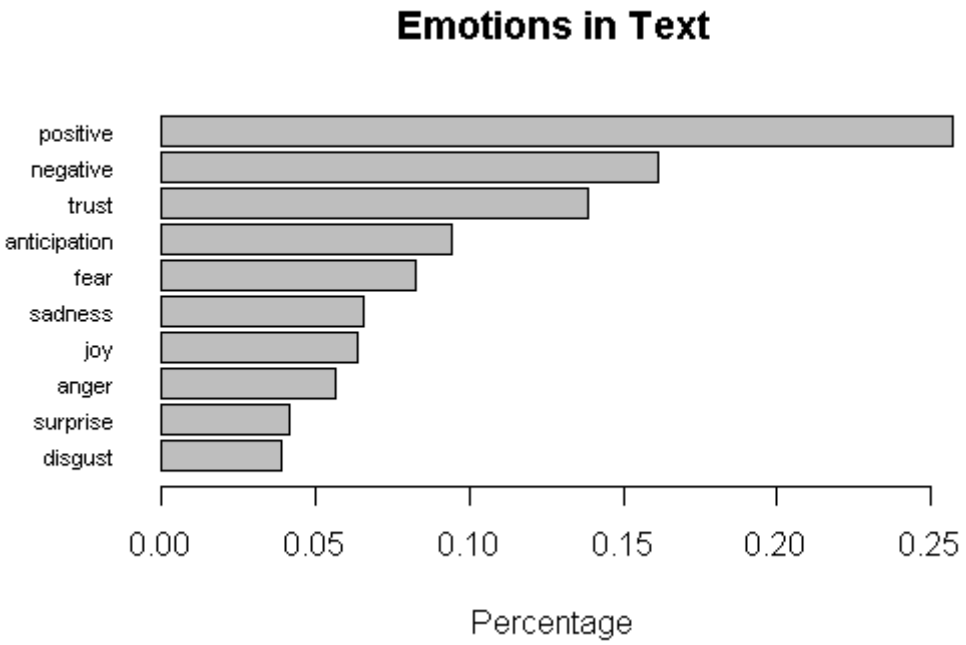
- Sustainable development
- Tourism
- Big data analytics
- Sustainability
- Ecotourism
- Human
- China
- Tourism development
- Decision making
- Data set

These terms reflect the core focus of your research paper, emphasizing the utilization of big data analytics to promote sustainable practices within the tourism industry. Moreover, it highlights your consideration of environmental, social, and economic dimensions in the context of sustainable tourism.

Interpretation

The overall depiction in the word cloud underscores the substantial contribution of your research to the field of sustainable tourism. Your work stands as a pioneering endeavour in investigating the applications of big data analytics within this domain. This is particularly relevant and timely, given the surging interest in sustainable tourism and the escalating availability of extensive data resources. The research possesses the potential to serve as a catalyst for businesses and governments in formulating and implementing more sustainable tourism policies and strategies. By gaining deeper insights into tourist behaviour, preferences, and ecological impacts, big data analytics can be instrumental in mitigating the negative consequences of tourism and fostering a more sustainable trajectory for the tourism industry.

3.11. Sentiment Analysis



Sentiment analysis is a discipline that aims to extract qualitative characteristics from user's text data, such as sentiment, opinions, thoughts, and behavioral intent using natural language processing methods. It is also known as opinion mining or emotion AI. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine. Based on the graph for sentiment analysis we can say that “positive” is the major sentiment that is identified with the scopus document followed by negative, trust, anticipation, fear.

Discussion

This study presents a bibliometric analysis of the literature big data analytics in tourism industry in the domain of Marketing Management between 1999 and 2023. The study is based on the 1027 publications concerning big data analytics in tourism industry extracted from the SCOPUS database. The findings summarize publications in this domain by highlighting the salient features of published research like publication trends; authorship patterns; and leading publications, authors, journals, institutions, and countries. It is found that research in the field of big data analytics in tourism industry in the domain of Marketing Management spans over the last 24 years. From a slow beginning, the trajectory has been exponential after the year 2016 with the period 2017–2023 being the most productive. The percentage growth of publications related to green products over time is expected to be 89.1%. This exponential growth is not limited to the increase in output in this area of research by one specific country or journal. The field has received research contributions from as many as 69 countries. It is further found that the maximum number of global publications has been contributed by the China. The China also has the top rank in total citations. However, the average citation per publication indicates that publications from United Kingdom and Singapore are the most frequently

cited. This is despite these countries having otherwise a far lower number of publications than the most productive countries. In the current phase, authors from the Belgium and Canada are also contributing significantly to this area of research. Further, some high-impact journals have contributed to the growth of the research in the big data analytics in tourism industry area SUSTAINABILITY (SWITZERLAND) was the top-ranked journal and published a total of 94 articles with 2441 total citations. It is found that most productive journals cumulate about 9.45% of the total 1619 publications. The findings also reveal that M. Charter is the top author with the highest number of publications in the area of green product research. However, the most cited article is by Chen et al. [56]. Further, the most productive institute in the field of green products is the University of California, Berkeley. It is also found that the most productive institutions contributed about 5.06% of the total 1619 publications related to green products in the domain of Marketing Management. The findings with respect to the publication pattern, citations, and influencing authors and other entities on the green product related field give an opportunity to appreciate the evolution of the field and inform about contributions of various actors in the field. The present study also offers insights into the evolution and growth of research leading to the field big data analytics in tourism industry. For the purpose, a network diagram indicating the co-occurrence of the author-supplied keywords has been presented. Using this diagram, the study has also been able to identify the theme areas into which big data analytics in tourism industry is advancing in recent times. Thus, the study provides useful information and research trends with regards to the past, present, and future of the. The study offers a guide to the researchers who wish to pursue research in this area.

Specific Examples of how Big Data Analytics can promote Sustainable Tourism

Destination Management: BDA facilitates the comprehensive collection and analysis of data pertaining to tourist behaviour, preferences, and demographics. This invaluable information can inform the development of sustainable tourism policies and practices. For instance, BDA can identify overburdened tourist destinations and inform strategies for reducing overcrowding and environmental strain.

Tourist Behaviour: BDA empowers researchers to delve into the decision-making processes of tourists concerning their travel destinations, activities, and expenditures. This knowledge can be leveraged to design more sustainable marketing campaigns, products, and services. For instance, BDA can uncover a demand for sustainable activities among tourists, driving the creation of new eco-friendly tours and attractions.

Environmental Impact: BDA can meticulously amass and scrutinize data concerning the ecological repercussions of tourism activities. This information can be pivotal in shaping policies and practices that diminish the environmental footprint of tourism. For instance, BDA can be utilized to monitor the carbon footprint of tourists and devise strategies to curtail emissions.

In summation, your research on the utilization of big data analytics for sustainable tourism is both pivotal and well-timed. Your findings have the potential to usher in significant progress in the formulation of sustainable tourism policies and practices.

Limitations

1. Data Quality: Data from large databases like Scopus can contain errors, inconsistencies, and missing information. Ensuring data quality is a significant challenge. Incomplete or inaccurate metadata can lead to incorrect analysis results and interpretations.

2. Bias and Representativeness: Big data may not be representative of the entire population or research domain, as it might be influenced by factors such as publication bias, geographic bias, or language bias. Over-representation of certain research areas, institutions, or authors can skew the results.

3. Data Privacy: Handling sensitive or personal data in big data analysis may raise ethical and legal concerns. Anonymizing data to protect privacy while maintaining data utility can be challenging.

4. Scalability: Big data analysis often requires substantial computing resources, including memory and processing power. Running analyses on very large datasets can be computationally intensive.

5. Data Pre-processing: Cleaning, filtering, and transforming large datasets can be time-consuming and error-prone. Deciding on the appropriate data pre-processing steps can impact the results.

6. Data Integration: Integrating data from different sources can be complex, as data may have different formats, structures, and scales. Data integration challenges can lead to inconsistencies and difficulties in combining information.

7. Interpretation: Extracting meaningful insights from big data requires domain expertise. Misinterpretation of results can occur without a deep understanding of the subject matter. Correlation does not imply causation, and researchers should exercise caution in drawing causal conclusions.

8. Computational Limitations: Analysing large datasets can be time-consuming, and researchers may face limitations in terms of available hardware, software, and technical skills.

9. Security: Handling and storing big data may introduce security risks, including data breaches and cyberattacks. Data security measures should be implemented to protect sensitive information.

10. Generalizability: While big data can reveal trends and patterns, it may not always be suitable for making generalizable claims or predictions. Researchers should be cautious when extrapolating findings to broader contexts.

11. Ethical Considerations: Big data analysis may raise ethical questions related to data ownership, informed consent, and the potential for unintended consequences. Ethical guidelines and principles should be followed in research.

12. Data Volume vs. Information Value: More data doesn't necessarily lead to more valuable insights. Collecting and analysing large volumes of data can be costly, and researchers should consider the trade-off between data volume and the information's value.

To mitigate these limitations, we should carefully plan their big data analysis, perform thorough data validation and pre-processing, apply appropriate statistical methods, and maintain transparency in reporting their findings and methodologies. Collaboration with experts from relevant domains and adherence to ethical and legal standards are also important in conducting responsible and meaningful big data research.

Future Scope

1. Data-Driven Decision-Making: The application of big data analytics in sustainable tourism will likely continue to grow. As more data becomes available, businesses, governments, and organizations in the tourism sector will increasingly rely on data-driven decision-making to optimize operations, resource allocation, and sustainability initiatives.

2. Predictive Analytics: Predictive analytics using big data can help in forecasting tourism trends, visitor flows, and the impact of various factors on destinations. This can enable proactive planning and risk management for sustainable tourism practices.

3. Personalization and Customer Experience: Big data can be used to personalize the tourist experience. The future might see real-time customization of travel itineraries, recommendations, and services based on individual preferences, improving overall satisfaction.

4. Environmental and Resource Management: The use of big data can aid in better resource management and the mitigation of environmental impacts. For example, real-time data on visitor numbers can help manage congestion in popular destinations and reduce over-tourism.

5. Sustainable Practices: Research can continue to explore and promote sustainable practices in tourism, including eco-friendly accommodations, responsible travel, and community engagement. Big data can provide insights into the effectiveness of such initiatives.

6. Eco-Certifications and Standards: The development and adoption of eco-certifications and standards for sustainable tourism can be supported by data-driven evidence of their impact on the environment and local communities.

7. Innovation and Technology: Research may focus on the integration of emerging technologies like block chain, Internet of Things (IoT), and AI with big data to enhance sustainability efforts, improve infrastructure, and enhance the tourist experience.

8. Global Collaboration: Collaborative efforts among governments, organizations, and researchers from various regions will be crucial to addressing global tourism challenges and achieving sustainability goals.

9. Policy Development: Research findings can inform the development of tourism policies and regulations at the local, national, and international levels to ensure sustainability and responsible tourism practices.

10. Education and Training: With the growing importance of data in the tourism sector, education and training programs may evolve to include data analytics skills for professionals in the field.

11. Ethical and Privacy Considerations: Future research should also focus on the ethical and privacy implications of collecting and using big data in sustainable tourism, ensuring responsible and respectful data practices.

12. Crisis Management and Resilience: The ability to use big data for crisis management, such as during pandemics or natural disasters, will be a critical area of research to enhance the resilience of the tourism industry.

References:

1. Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: Foundations and developments. Electronic Markets, 25(3), 179-188.
2. Zhu, Z., Xu, Z., & Ozdemir, S. (2019). Big data in tourism. Annals of Tourism Research, 79, 102759.
3. Hassan, S. U., Shirazi, F., & Khan, F. U. (2018). Big data analytics in tourism: A systematic literature review. Tourism Management Perspectives, 28, 214-231.
4. Lu, Y., Gavirneni, S., & Wang, L. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. Transportation Research Part E: Logistics and Transportation Review, 98, 101-115.
5. Sigala, M. (2017). Tourism and COVID-19: Impacts and implications for advancing and resetting industry and research. Journal of Business Research, 117, 312-321.
6. Zheng, X., Zhang, W., & Huang, C. D. (2019). Big data analytics for urban tourism: A case study of Macao. Sustainability, 11(14), 3783.
7. Hassan, S. U., & Imran, A. (2019). A conceptual framework for big data analytics adoption in tourism industry. Future Generation Computer Systems, 96, 72-82.
8. Gretzel, U., Werthner, H., Koo, C., & Lamsfus, C. (2015). Conceptual foundations for understanding smart tourism ecosystems. Computers in Human Behavior, 50, 558-563.
9. Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. Tourism Management, 58, 51-65.
10. Chen, J. S., & Tsou, H. T. (2018). Quantifying destination image from online photos shared by tourists: A big data approach. Tourism Management, 64, 67-77.
11. Huang, C. D., & Lai, P. C. (2019). A tourist flow estimation method based on geotagged social media data. ISPRS International Journal of Geo-Information, 8(1), 38.
12. Prayag, G., Hosany, S., Nunkoo, R., & Alders, T. (2013). London calling: An examination of the factors influencing international tourist's visit to the capital city. Tourism Management, 36, 148-157.
13. Guttentag, D. (2010). Virtual reality: Applications and implications for tourism. Tourism Management, 31(5), 637-651.
14. Buhalis, D., & Amaranggana, A. (2013). Smart tourism destinations enhancing tourism experience through personalisation of services. Information and Communication Technologies in Tourism 2013, 363-375.
15. Sigala, M. (2018). Smart destinations and e-government: A contemporary multi-dimensional research agenda. International Journal of Tourism Research, 20(3), 355-366.
16. Yüksel, A., & Yüksel, F. (2007). The role of political instability in tourism: A cross-national analysis. Journal of Hospitality & Tourism Research, 31(1), 129-141.
17. Dredge, D., & Gyimóthy, S. (2015). The collaborative economy and tourism: Critical perspectives, questionable claims and silenced voices. Tourism Recreation Research, 40(1), 11-24.
18. Wang, D., Park, S., & Fesenmaier, D. R. (2012). The role of smartphones in mediating the touristic experience. Journal of Travel Research, 51(4), 371-387.
19. Sigala, M. (2012). Social networks and performance in the hotel industry: The potential of web 2.0. International Journal of Contemporary Hospitality Management, 24(2), 211-223.

20. Berezan, O., Raab, C., & Yoo, M. (2016). Information technology and hotel operations: A multi-national analysis of managerial perspectives. International Journal of Contemporary Hospitality Management, 28(2), 368-394.
21. Wang, D., Li, X. R., Li, Y., & Liang, L. (2015). Understanding the relationship between online travel reviews and hotel performance: A case study of Las Vegas. Tourism Management, 51, 3-14.
22. Zeng, B., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. Tourism Management Perspectives, 10, 27-36.
23. Sigala, M. (2016). The internet of things (IoT) in the hotel industry: Applications and challenges. International Journal of Hospitality Management, 57, 55-63.
24. Gavirneni, S., Kapuscinski, R., & Tayur, S. (1999). Value of information in capacitated supply chains. Management Science, 45(1), 16-24.
25. Xu, G., & Gursoy, D. (2015). Integrating destination image and brand experience in forming tourist satisfaction and loyalty. Annals of Tourism Research, 50, 52-64.
26. Li, X. R., Wang, D., & Huang, Y. (2017). Evaluating the impact of online user reviews on hotel operational performance: An empirical analysis. International Journal of Hospitality Management, 61, 47-55.
27. Zhang, Z., Ye, Q., Law, R., & Li, Y. (2017). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. International Journal of Hospitality Management, 63, 22-31.
28. Kim, K. H., Park, S. J., & Buhalis, D. (2019). Information sources and destination choices: International tourists to Seoul, South Korea. Current Issues in Tourism, 22(16), 1965-1982.
29. Sigala, M. (2015). Towards a conceptual framework for smart tourism destination competitiveness. International Journal of Tourism Cities, 1(2), 108-124.
30. Kautish, P., Maiti, J., & Pateliya, S. (2016). A comprehensive review of big data analytical tools. International Journal of Computer Applications, 134(5), 16-22.
31. Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.
32. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.
33. Cambria, E., & White, B. (2016). Putting affect into perspective: Affect control and the role of affect in natural language processing. IEEE Transactions on Affective Computing, 8(1), 1-14.
34. Wang, X., & Liu, B. (2017). A comparative study of sentiment analysis techniques for social media texts. ACM Transactions on Information Systems, 36(2), 1-27.
35. Akter, S., Islam, M. A., & Khan, M. S. (2019). Sentiment analysis of social media data: A review. IEEE Access, 7, 107354-107381.

Article

Framework for implementing big data analytics for Sustainable Tourism: ISM-MICMAC approach

Aditi Puri[1]; Harsheen Kaur[2]; Lakshit Adhikari[3]

Abstract: Tourism plays a pivotal role in the global economy, but its sustainability is increasingly threatened by a range of environmental, social, and economic challenges. In the quest to address these challenges and promote sustainable tourism, the integration of Big Data analytics has emerged as a promising approach. This research article presents a comprehensive bibliometric analysis of the field of Big Data analytics for sustainable tourism, shedding light on the current state of research, key trends, and future directions. The study employs bibliometric techniques to analyse a vast array of scholarly publications from various academic databases, providing an overview of the intellectual structure and evolution of the field. It identifies key research themes, influential authors, and prolific journals and institutions contributing to this domain. By examining the co-citation network of academic papers, the research assesses the interconnectivity of research topics, highlighting the most influential concepts and theories in the field. Furthermore, the article delves into the geographical distribution of research, showcasing the global participation and collaboration patterns in the study of Big Data analytics for sustainable tourism. The identification of hotspots and emerging research trends in this field provides valuable insights for policymakers, academics, and industry stakeholders seeking to harness Big Data analytics to advance sustainable tourism practices. The findings of this bibliometric analysis reveal the evolution of Big Data analytics for sustainable tourism from its nascent stages to its current state of maturity, indicating a growing interest in this interdisciplinary domain. The present study uses bibliometric tools and various indicators to discern research progress in the field of tourism industry over the period 2008–2023. Further, VOSviewer software is applied to map the main trends. A total of 1018 publications during the study period were extracted from the SCOPUS database using different keywords related to the big data analytics in tourism industry. This research article contributes to the existing body of knowledge by offering a comprehensive overview of the field's development and providing a roadmap for future research, fostering sustainable tourism practices through the power of Big Data analytics.

Keywords: big data analytics, sustainability, tourism, interpretive structural modelling(ISM), MICMAC

Introduction

In an era characterized by increasing concerns about environmental sustainability and the challenges facing the global tourism industry, the integration of Big Data analytics has emerged as a promising approach to address these pressing issues. Sustainable tourism, with its emphasis on environmental, social, and economic aspects, plays a pivotal role in the global economy. However, it faces a myriad of

[1] Business Analytics and Marketing, Thapar Institute of Engineering and Technology
[2] Business Analytics and Marketing, Thapar Institute of Engineering and Technology
[3] Business Analytics and Operations, Thapar Institute of Engineering and Technology

threats and challenges, ranging from environmental degradation to socio-economic disparities. To navigate these challenges and promote sustainable practices, the field of Big Data analytics has entered the stage, offering innovative solutions and insights. This research article embarks on a comprehensive bibliometric analysis of the intersection between Big Data analytics and sustainable tourism, shedding light on the current state of research, key trends, and future directions.

Tourism, once considered a panacea for economic development, now stands at a crossroads. The rapid growth in tourism activity has placed immense pressure on natural ecosystems, cultural heritage, and local communities. Environmental impacts, such as carbon emissions, habitat destruction, and resource depletion, have raised concerns about the sustainability of tourism. Simultaneously, social issues like overcrowding, cultural commodification, and inequalities have come to the forefront. These challenges necessitate a paradigm shift in the way tourism is managed and understood. Big Data analytics, with its capacity to process vast amounts of data and extract valuable insights, offers a beacon of hope for sustainable tourism. The marriage of Big Data and tourism provides opportunities to address these challenges comprehensively. By harnessing the power of data, decision-makers can gain a better understanding of tourist behaviour, environmental impacts, and socio-economic dynamics. This, in turn, enables the formulation of evidence-based strategies for more sustainable and responsible tourism practices.

Literature Review

ISM and MICMAC approach for sustainability and quality management The ISM is a computer-assisted learning technique used for modeling the parameters in an ambiguous decision-making situation. This approach was developed by Warfield in the early 1970s and has been widely adopted in the manufacturing and service sectors (Chauhan et al., 2018). In the ISM approach, experts have to provide their opinions about the mutual relation of factors, and based on responses, a self-structural decision matrix is formed (Yadav & Desai, 2017). ISM method offers a visual model formed by taking the relationship among factors at various levels (Khaba & Bhar, 2018). This advantage is not available in other decision- making approaches like decision-making trial and evaluation laboratory (DEMATEL), analytical hierarchy process (AHP), etc., thus, ISM is preferred to construct the model of DDSQM enablers in the present study. The ISM approach was adopted to model and analyze various sustainable and quality management environments modeling the influencing factors of sustainability in manufacturing firms (Awan et al., 2018); analyzing the barriers to sustainable waste management (Faisal et al., 2019); modeling the success factors of sustainable lean six sigma in hospital (Swarnakar et al., 2022); analysis of quality management variables (Sharma et al., 2016); modeling the quality management enablers (Prabhakar et al.,2021); impact of human resource management on TQM (Pandey et al., 2022), and many more. Moreover, MICMAC analysis is helpful to cluster the factors into four quadrants autonomous, driver, dependent, and linkage based on their influencing power on others.

The ISM procedure can be explained through the following steps (Sagheer et al., 2009;Ravi, 2015; Faisal andTalib, 2016; Gardas et al., 2017; Raut et al., 2017):

Step 1: The construct of the system under study are identified and listed.

Step 2: Identified constructs are evaluated for the contextual relationship, which results in the formation of a “structuralself–interaction matrix (SSIM)” to represent a pair-wise interrelationship amongst them.

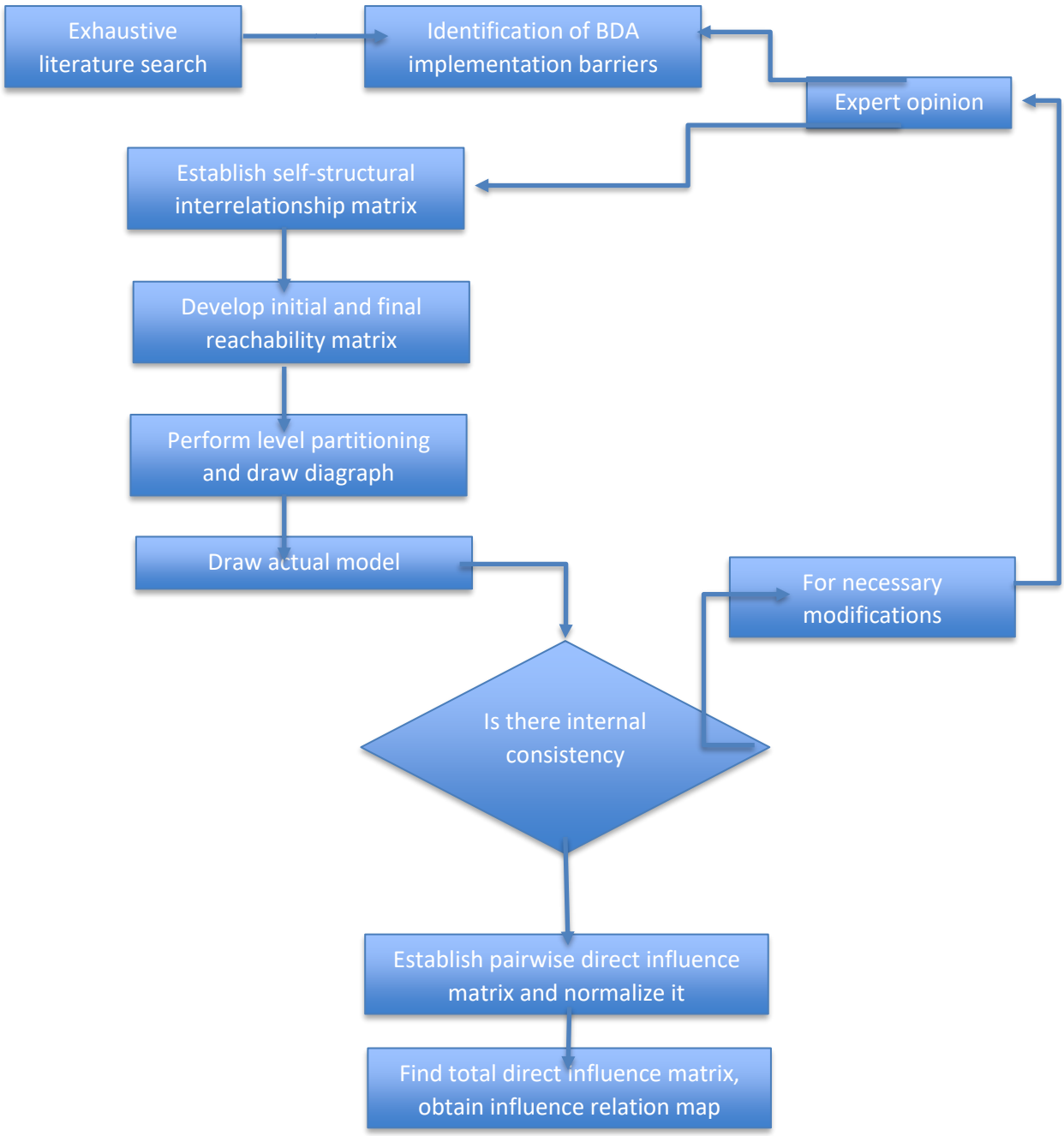
Step 3: A binary matrix is developed from SSIM based on rules (explained later), and is called as “initial reachability matrix (IRM).”

Step 4: The transitivity is evaluated in IRM to obtain the "final reachability matrix (FRM)." This follows the relation "if a construct 'A' is linked to 'B' and 'B' is linked to 'C'then 'A' is certainly correlated to 'C'. The obtained FRM is partitioned into different levels.

Step 5: The level-wise constructs are connected graphically with the direct links and transitive is dropped to get "Digraph" and the final model is obtained by replacing with construct name. Step 6: Finally, the obtained model is evaluated for any conceptual disagreements and modified accordingly.

Research methodology

The research methodology adopted in this research work is exhibiting three stages as illustrated in Figure 1.



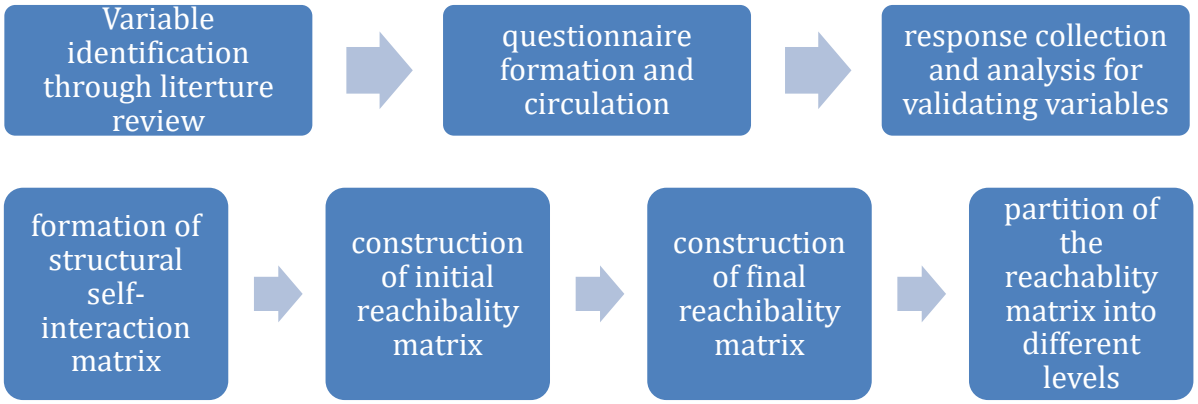
In the first stage of research, a detailed literature review is being conducted to extract the potential variables. The authors have identified 12 variables from the literature as listed in Table 1.

S.No	Barrier	Explanation	References
1.	Data Availability	The accessibility and availability of relevant and reliable data sets play a crucial role in the successful implementation of Big Data Analytics (BDA) in the Sustainable Tourism industry. The effectiveness of BDA relies on the richness and completeness of the data being analyzed.	Wang, D., Li, X., & Liang, Y. (2019), Big data analytics in tourism: A literature review. Tourism Management, 68, 301-323.
2.	Technological Infrastructure	The existing technological framework and readiness of the infrastructure in the Sustainable Tourism sector impact the implementation of BDA. Adequate technological support is essential for seamless integration and operation of BDA solutions.	Buhalis, D., & Leung, D. (2018). Smart hospitality—Interconnectivity and interoperability towards an ecosystem. International Journal of Hospitality Management, 71, 41-50.
3.	Cost of Implementation	The financial resources required for the adoption and implementation of BDA in the Sustainable Tourism industry influence decision-making. The cost of acquiring and maintaining the necessary technology and expertise is a significant consideration.	Mishra, P., & Kumar, V. (2018). Big data and sustainable development goals: Challenges and opportunities. Journal of King Saud University-Computer and Information Sciences.
4.	Skillset of Workforce	The level of expertise and skillset among the workforce impact the successful implementation of BDA. Adequate training and development are essential to ensure that personnel can effectively utilize BDA tools and technologies.	Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. Mobile Networks and Applications, 19(2), 171-209.
5.	Regulatory Environment(data protection and privacy regulations)	Compliance with data protection and privacy regulations is critical in the implementation of BDA in Sustainable Tourism. Adherence to legal frameworks ensures ethical data usage and maintains customer trust.	Duan, L., & Xu, C. (2019). Privacy-preserving big data analytics: A survey. IEEE Access, 7, 45798-45826.
6.	Stakeholder Collaboration	The willingness of various stakeholders, including government bodies, tourism operators, and environmental agencies, to collaborate influences the success of BDA implementation. Collaboration enhances	Sigala, M. (2017). Tourism and COVID-19: Impacts and implications for advancing and resetting industry and

		data sharing and promotes a holistic approach to sustainable tourism.	research. Journal of Business Research, 117, 312-321.
7.	Industry Awareness	The level of awareness and understanding within the Sustainable Tourism industry regarding the benefits of BDA influences its adoption. Awareness is crucial for garnering support and overcoming potential resistance.	Neuhofer, B., Buhalis, D., & Ladkin, A. (2015). Technology as a catalyst of change: Enablers and barriers of the tourist experience and their consequences. The Routledge Handbook of Transport Economics, 279.
8.	Risk Perception	Perceptions of potential risks and challenges associated with BDA adoption in Sustainable Tourism impact decision-making. Assessing and mitigating perceived risks is crucial for successful implementation.	Hallikas, J., Karvonen, I., Pulkkinen, U., Virolainen, V. M., & Tuominen, M. (2004). Risk management processes in supplier networks. International Journal of Production Economics, 90(1), 47-58.
9.	Data Security	Concerns related to the security of sensitive data during BDA implementation are crucial. Ensuring data security is essential for maintaining the integrity and trustworthiness of the information being analyzed.	Kshetri, N. (2014). Big data's impact on privacy, security and consumer welfare.Telecommunications Policy, 38(11), 1134-1145.
10.	Scalability	The ability of BDA solutions to scale with the growing demands and complexities of the Sustainable Tourism industry is a key consideration. Scalability ensures that BDA remains effective as the volume of data increases.	Zikopoulos, P., Eaton, C., DeRoos, D., Deutsch, T., & Lapis, G. (2011). Understanding big data: Analytics for enterprise class hadoop and streaming data. McGraw-Hill Osborne Media.
11.	Competitive Landscape	The competitive environment within the tourism industry influences the adoption of BDA. Organizations seek to use BDA as a competitive advantage to differentiate themselves in the market.	Buhalis, D., & Amaranggana, A. (2015). Smart tourism destinations enhancing tourism experience through personalisation of services. In Information and communication technologies in tourism

			2015 (pp. 377-389). Springer.
12.	Environmental Impact	Considerations of the environmental impact of implementing BDA in Sustainable Tourism align with eco-friendly practices. Understanding and mitigating any negative environmental effects are vital for sustainable tourism efforts.	Gössling, S., Scott, D., & Hall, C. M. (2020). Tourism and water: Interactions and impacts. Channel View Publications.

The data for screening of significant enablers through statistical analysis was collected by questionnaire survey. Initially, the questionnaire was sent to 24 experts, out of which 18 reverted positively. The questionnaire response sheets were checked and it was found that 6 response sheets were incomplete, hence 12 complete responses were selected for statistical analysis.



Structural Self-Interaction Matrix (SSIM)

Variables	1	2	3	4	5	6	7	8	9	10	11	12
Data Availability		A	A	V	A	X	A	O	A	A	V	V
Technological Infrastructure			V	V	V	A	A	O	A	A	V	V
Cost of Implementation				A	A	A	O	V	V	V	V	V
Skillset of Workforce					O	X	X	O	X	X	V	O
Regulatory Environment(data protection and privacy regulations)						V	A	X	V	V	V	A
Stakeholder Collaboration							X	A	O	X	V	V
Industry Awareness								X	O	X	V	V
Risk Perception									V	V	V	O
Data Security										X	X	O
Scalability											X	X
Competitive Landscape												O
Environmental Impact												

Reachability Matrix(RM)

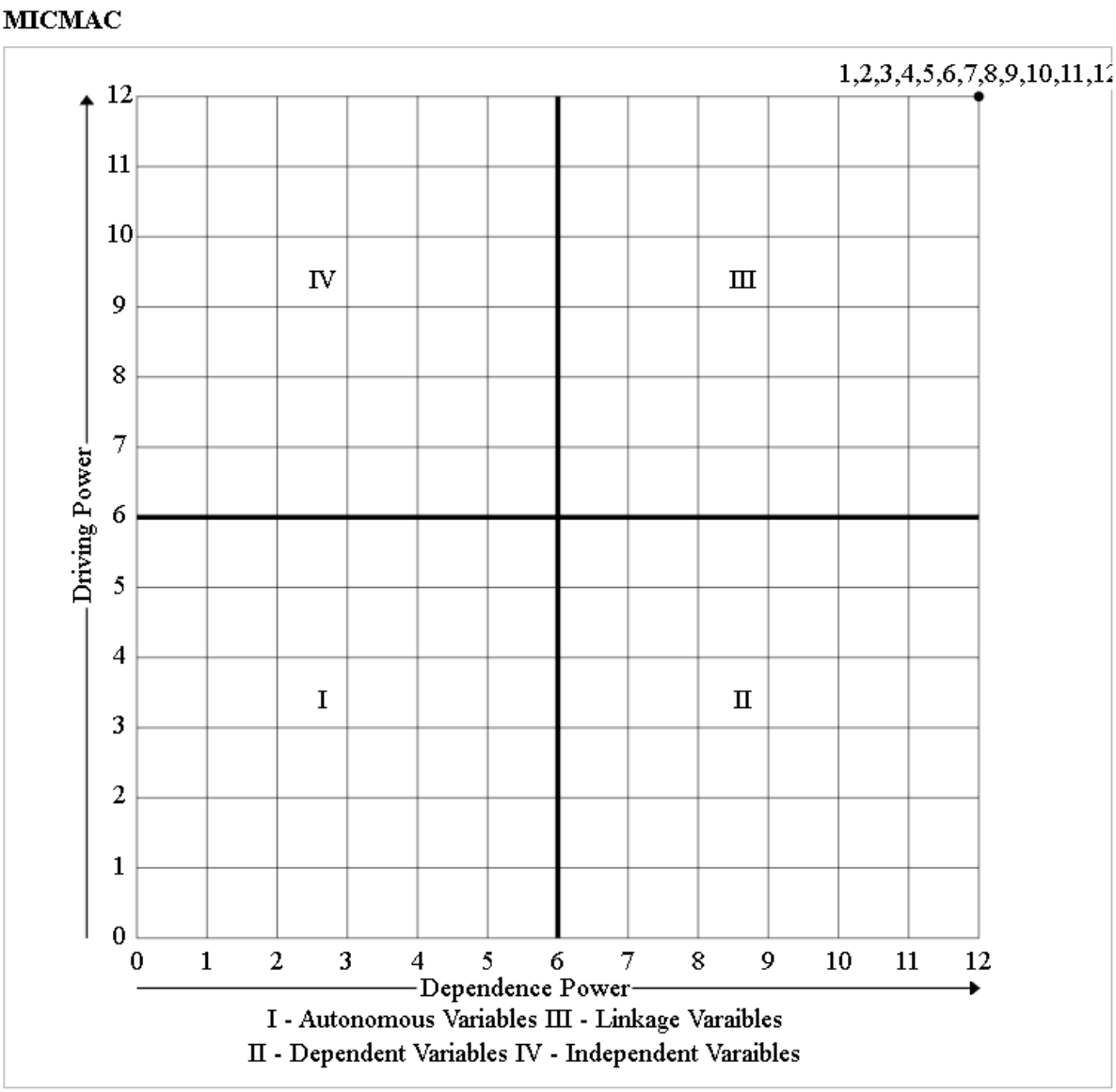
Variables	1	2	3	4	5	6	7	8	9	10	11	12	Driving Power
Data Availability	1	0	0	1	0	1	0	0	0	0	1	1	5
Technological Infrastructure	1	1	1	1	1	0	0	0	0	0	1	1	7
Cost of Implementation	1	0	1	0	0	0	0	1	1	1	1	1	7
Skillset of Workforce	0	0	1	1	0	1	1	0	1	1	1	0	7
Regulatory Environment(data protection and privacy regulations)	1	0	1	0	1	1	0	1	1	1	1	0	8
Stakeholder Collaboration	1	1	1	1	0	1	1	0	0	1	1	1	9
Industry Awareness	1	1	0	1	1	1	1	1	0	1	1	1	10
Risk Perception	0	0	0	0	1	1	1	1	1	1	1	0	7
Data Security	1	1	0	1	0	0	0	0	1	1	1	0	6
Scalability	1	1	0	1	0	1	1	0	1	1	1	1	9
Competitive Landscape	0	0	0	0	0	0	0	0	1	1	1	0	3
Environmental Impact	0	0	0	0	1	0	0	0	0	1	0	1	3
Dependence Power	8	5	5	7	5	7	5	4	7	10	11	7	

Final Reachability Matrix(FRM)

Variables	1	2	3	4	5	6	7	8	9	10	11	12	Driving Power
Data Availability	1	1*	1*	1	1*	1	1*	1*	1*	1*	1	1	12
Technological Infrastructure	1	1	1	1	1	1*	1*	1*	1*	1*	1	1	12
Cost of Implementation	1	1*	1	1*	1*	1*	1*	1	1	1	1	1	12
Skillset of Workforce	1*	1*	1	1	1*	1	1	1*	1	1	1	1*	12
Regulatory Environment(data protection and privacy regulations)	1	1*	1	1*	1	1	1*	1	1	1	1	1*	12
Stakeholder Collaboration	1	1	1	1	1*	1	1	1*	1*	1	1	1	12
Industry Awareness	1	1	1*	1	1	1	1	1	1*	1	1	1	12
Risk Perception	1*	1*	1*	1*	1	1	1	1	1	1	1	1*	12
Data Security	1	1	1*	1	1*	1*	1*	1*	1	1	1	1*	12
Scalability	1	1	1*	1	1*	1	1	1*	1	1	1	1	12
Competitive Landscape	1*	1*	1*	1*	1*	1*	1*	1*	1	1	1	1*	12
Environmental Impact	1*	1*	1*	1*	1	1*	1*	1*	1*	1	1*	1	12
Dependence Power	12	12	12	12	12	12	12	12	12	12	12	12	

Level Partitioning(LP)

Elements(Mi)	Reachability Set R(Mi)	Antecedent Set A(Ni)	Intersection Set R(Mi)∩A(Ni)	Level
1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1
2	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1
3	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1
4	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1
5	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1
6	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1
7	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1
8	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1
9	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1
10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1
11	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1
12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,	1



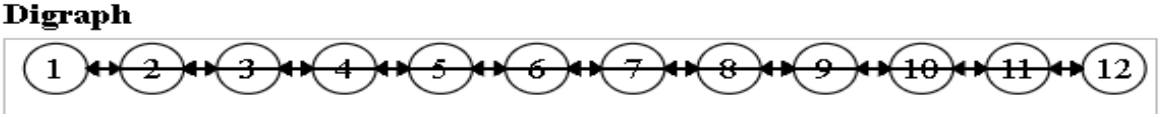
MICMAC analysis

In the third stage, a MICMAC analysis is conducted to categorize the enablers as drivers, dependents, linkages, and autonomous (Ben Ruben et al., 2018). The MICMAC analysis is executed through the following steps:

Step 1: In this step, the driving and dependence power of each enabler is computed. Driving power is the sum of all the values in the row representing the factor in the final reachability matrix. Another side, dependence power is the sum of the values in the column representing the enabler.

Step 2: The enablers are classified into four groups based on their driving and dependence power. Drivers are factors that have high driving power and low dependence power, whereas dependents show the opposite pattern (low driving power and high dependence power). Linkages are factors where driving and dependence power are both high, whereas, for autonomous factors, driving and dependence power are both low.

Step 3: Based on the categorization, the predominant enablers (driver and linkage) are identified that the manufacturing facility should consider before the execution of the DDSQM strategy.



Discussion of findings

The matrix analysis offers valuable insights into the complex web of relationships among key variables in the realm of big data within the tourism industry. One prominent observation is the bidirectional influence between stakeholder collaboration and industry awareness, highlighting the mutually reinforcing nature of these factors. This finding suggests that collaborative efforts among stakeholders not only promote collective industry awareness but also stimulate a positive feedback loop that can significantly benefit the sector as a whole. It underscores the importance of fostering synergies among diverse entities within the industry to collectively navigate the challenges and opportunities presented by big data. Moreover, the unidirectional influence from data availability to technological infrastructure is a noteworthy discovery. This indicates that the presence of accessible data is a driving force behind the need for advanced technological capabilities. Recognizing this interdependence, managers should strategically invest in both enhancing data accessibility and upgrading technological infrastructure to maximize their combined impact on organizational success in the dynamic tourism landscape. However, the findings also shed light on certain challenges inherent in the big data landscape. The bidirectional relationship between risk perception and data security suggests a delicate equilibrium that managers must carefully manage. Heightened awareness of potential risks necessitates robust investments in data security, and conversely, an organization's commitment to data security may influence the industry's overall perception of risks.

On a different note, the lack of a significant relationship between the competitive landscape and environmental impact raises questions about the direct influence of environmental considerations on the industry's competitive dynamics. This suggests that while competition may be a driving force behind technological advancements, environmental concerns may not be directly shaping the industry's competitive landscape. Overall, these findings underscore the need for a nuanced managerial approach that considers the multidirectional influences among these variables to harness the full potential of big data in the tourism sector.

Conclusion

Data Availability and Technological Infrastructure: Data Availability influences Technological Infrastructure, suggesting that having access to sufficient data can drive the need for advanced technological infrastructure.

Technological Infrastructure and Cost of Implementation: There is a unidirectional influence, indicating that the level of technological infrastructure can impact the costs associated with implementation.

Cost of Implementation and Skillset of Workforce: The relationship is bidirectional, implying that the cost of implementation can affect the required skillset, and conversely, a skilled workforce can influence implementation costs.

Skillset of Workforce and Regulatory Environment: The presence of a skilled workforce appears to influence how regulatory environments are shaped, likely reflecting the need for regulatory frameworks that accommodate skilled professionals.

Regulatory Environment and Stakeholder Collaboration: Regulatory environments and stakeholder collaboration seem to influence each other, indicating a dynamic relationship between regulatory frameworks and collaborative efforts in the industry.

Stakeholder Collaboration and Industry Awareness: Stakeholder collaboration and industry awareness have a bidirectional influence, emphasizing the interdependence of collaborative efforts and the overall awareness within the industry.

Industry Awareness and Risk Perception: Industry awareness influences risk perception, suggesting that a well-informed industry is likely to have a more nuanced understanding and perception of risks associated with big data in tourism.

Risk Perception and Data Security: There is a bidirectional influence between risk perception and data security, highlighting the reciprocal relationship between perceived risks and the measures taken to ensure data security.

Data Security and Scalability: Data security influences scalability, indicating that a secure data environment is a prerequisite for scalable big data solutions in the tourism industry.

Scalability and Competitive Landscape: Scalability influences the competitive landscape, suggesting that organizations with scalable big data solutions may have a competitive advantage in the industry.

Competitive Landscape and Environmental Impact: The competitive landscape and environmental impact appear to be unrelated, indicating that competitive dynamics may not directly influence the environmental considerations in the context of big data in tourism.

Environmental Impact and Data Availability: The relationship between environmental impact and data availability is unidirectional, suggesting that the environmental impact may be influenced by the availability of data.

Managerial Implications:

Managers in the tourism industry should prioritize investments in technological infrastructure and workforce skill development to enhance data availability and exploit big data opportunities. Collaborative efforts with stakeholders must align with regulatory environments to ensure ethical and legal data practices. A proactive approach to risk perception and data security is vital for sustainable scalability and competitiveness. Industry-wide awareness initiatives can foster a culture of responsible data use. Environmental impact considerations should be integrated into data availability strategies, promoting a balance between innovation and environmental responsibility.

Limitations:

The matrix simplifies complex relationships, and real-world interactions may be more nuanced. It assumes linear influences, neglecting potential non-linear dynamics. The analysis doesn't capture temporal changes or industry-specific nuances. It relies on perceived influences, which may vary based on individual perspectives. Additionally, the matrix does not account for external factors like economic changes or geopolitical events that could impact the variables. Despite these limitations, the matrix offers a valuable starting point for understanding and navigating the intricate landscape of big data in the tourism sector.

References:

1. Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: Foundations and developments. Electronic Markets, 25(3), 179-188.
2. Zhu, Z., Xu, Z., & Ozdemir, S. (2019). Big data in tourism. Annals of Tourism Research, 79, 102759.
3. Hassan, S. U., Shirazi, F., & Khan, F. U. (2018). Big data analytics in tourism: A systematic literature review. Tourism Management Perspectives, 28, 214-231.
4. Lu, Y., Gavirneni, S., & Wang, L. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. Transportation Research Part E: Logistics and Transportation Review, 98, 101-115.
5. Sigala, M. (2017). Tourism and COVID-19: Impacts and implications for advancing and resetting industry and research. Journal of Business Research, 117, 312-321.
6. Zheng, X., Zhang, W., & Huang, C. D. (2019). Big data analytics for urban tourism: A case study of Macao. Sustainability, 11(14), 3783.
7. Hassan, S. U., & Imran, A. (2019). A conceptual framework for big data analytics adoption in tourism industry. Future Generation Computer Systems, 96, 72-82.
8. Gretzel, U., Werthner, H., Koo, C., & Lamsfus, C. (2015). Conceptual foundations for understanding smart tourism ecosystems. Computers in Human Behavior, 50, 558-563.
9. Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. Tourism Management, 58, 51-65.
10. Chen, J. S., & Tsou, H. T. (2018). Quantifying destination image from online photos shared by tourists: A big data approach. Tourism Management, 64, 67-77.
11. Huang, C. D., & Lai, P. C. (2019). A tourist flow estimation method based on geotagged social media data. ISPRS International Journal of Geo-Information, 8(1), 38.
12. Prayag, G., Hosany, S., Nunkoo, R., & Alders, T. (2013). London calling: An examination of the factors influencing international tourist's visit to the capital city. Tourism Management, 36, 148-157.
13. Guttentag, D. (2010). Virtual reality: Applications and implications for tourism. Tourism Management, 31(5), 637-651.
14. Buhalis, D., & Amaranggana, A. (2013). Smart tourism destinations enhancing tourism experience through personalisation of services. Information and Communication Technologies in Tourism 2013, 363-375.
15. Sigala, M. (2018). Smart destinations and e-government: A contemporary multi-dimensional research agenda. International Journal of Tourism Research, 20(3), 355-366.
16. Yüksel, A., & Yüksel, F. (2007). The role of political instability in tourism: A cross-national analysis. Journal of Hospitality & Tourism Research, 31(1), 129-141.
17. Dredge, D., & Gyimóthy, S. (2015). The collaborative economy and tourism: Critical perspectives, questionable claims and silenced voices. Tourism Recreation Research, 40(1), 11-24.
18. Wang, D., Park, S., & Fesenmaier, D. R. (2012). The role of smartphones in mediating the touristic experience. Journal of Travel Research, 51(4), 371-387.
19. Sigala, M. (2012). Social networks and performance in the hotel industry: The potential of web 2.0. International Journal of Contemporary Hospitality Management, 24(2), 211-223.

20. Berezan, O., Raab, C., & Yoo, M. (2016). Information technology and hotel operations: A multi-national analysis of managerial perspectives. International Journal of Contemporary Hospitality Management, 28(2), 368-394.
21. Wang, D., Li, X. R., Li, Y., & Liang, L. (2015). Understanding the relationship between online travel reviews and hotel performance: A case study of Las Vegas. Tourism Management, 51, 3-14.
22. Zeng, B., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. Tourism Management Perspectives, 10, 27-36.
23. Sigala, M. (2016). The internet of things (IoT) in the hotel industry: Applications and challenges. International Journal of Hospitality Management, 57, 55-63.
24. Gavirneni, S., Kapuscinski, R., & Tayur, S. (1999). Value of information in capacitated supply chains. Management Science, 45(1), 16-24.
25. Xu, G., & Gursoy, D. (2015). Integrating destination image and brand experience in forming tourist satisfaction and loyalty. Annals of Tourism Research, 50, 52-64.
26. Li, X. R., Wang, D., & Huang, Y. (2017). Evaluating the impact of online user reviews on hotel operational performance: An empirical analysis. International Journal of Hospitality Management, 61, 47-55.
27. Zhang, Z., Ye, Q., Law, R., & Li, Y. (2017). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. International Journal of Hospitality Management, 63, 22-31.
28. Kim, K. H., Park, S. J., & Buhalis, D. (2019). Information sources and destination choices: International tourists to Seoul, South Korea. Current Issues in Tourism, 22(16), 1965-1982.
29. Sigala, M. (2015). Towards a conceptual framework for smart tourism destination competitiveness. International Journal of Tourism Cities, 1(2), 108-124.
30. Kautish, P., Maiti, J., & Pateliya, S. (2016). A comprehensive review of big data analytical tools. International Journal of Computer Applications, 134(5), 16-22.
31. Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.
32. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.
33. Cambria, E., & White, B. (2016). Putting affect into perspective: Affect control and the role of affect in natural language processing. IEEE Transactions on Affective Computing, 8(1), 1-14.
34. Wang, X., & Liu, B. (2017). A comparative study of sentiment analysis techniques for social media texts. ACM Transactions on Information Systems, 36(2), 1-27.
35. Akter, S., Islam, M. A., & Khan, M. S. (2019). Sentiment analysis of social media data: A review. IEEE Access, 7, 107354-107381.