Identifying high-value keywords for Bali tourism: A machine learning approach

Asep Koswara[1]

IKOPIN University, Jatinangor, INDONESIA[1]
[1]Email: aspkosw@gmail.com

Abstract – This study explores how machine learning techniques, specifically K-Means Clustering, can be applied to identify high-value tourism-related keywords for Bali using Google Keyword Planner data. By analyzing normalized search volume, cost-per-click (CPC), and competition scores, the keywords were grouped into three meaningful clusters: (1) high-volume, high-CPC keywords such as Bali luxury resorts and honeymoon in Bali, (2) low-volume, high-CPC keywords like eco-retreat Bali and private villa Ubud, and (3) high-volume, low-CPC keywords including Bali itinerary and Bali beach names. The Elbow Method confirmed that three clusters offered the optimal segmentation. Findings show that luxury and culturally immersive travel themes yield the highest marketing value, while niche keywords offer untapped advertising potential. Seasonal analysis revealed peaks in keyword interest during mid-year and year-end holiday seasons, aligning with international travel patterns. The machine learning approach enhanced keyword structuring and revealed strategic timing opportunities for content deployment. The study concludes that data-driven keyword targeting can significantly improve the effectiveness and sustainability of digital marketing in Bali's tourism sector. Recommendations include focusing on premium and niche clusters, while future research may integrate social media trends and geospatial analysis to deepen keyword relevance and behavioral insights.

Keywords: Bali Tourism; Keyword Analysis; Machine Learning; Tourism Marketing; Google Keyword Planner

1. Introduction

Tourism remains one of the most dynamic and economically significant sectors in Indonesia, contributing to national revenue, employment, and regional development. Among Indonesia's many destinations, Bali has long stood out as a global icon for culture, natural beauty, and spiritual retreats. In recent decades, Bali's tourism industry has undergone rapid transformation, influenced by technological advancements, shifting traveler behavior, and increasing competition in the digital marketplace (Suryani et al., 2023). As travelers increasingly rely on online platforms to plan and book their trips, search engine visibility has become a critical determinant of destination competitiveness. In this context, the identification of high-value keywords becomes a strategic priority for tourism stakeholders aiming to boost online reach, engage relevant audiences, and convert interest into actual visits.

The digital shift in tourism marketing has prompted researchers and practitioners alike to explore data-driven approaches for understanding traveler intent and behavior. One of the most powerful tools in this domain is Google Keyword Planner, which offers insights into search volume trends, keyword competition, and related search queries. By leveraging such data, destination marketing organizations (DMOs) can craft more targeted content and align their

online presence with user demand. However, given the sheer volume and complexity of keyword data, manual analysis becomes impractical. This has led to the growing adoption of machine learning (ML) techniques to extract actionable insights from large datasets (Núñez et al., 2024; Li et al., 2021).

In the context of Bali, recent studies highlight the importance of integrating technology into tourism planning and promotion. For instance, Ustriyana et al. (2023) applied random forest models to map agro-tourism potential across the island, demonstrating the feasibility and value of ML-based spatial analysis. Similarly, Koswara and Marpuah (2024) proposed an AI-driven SEO marketing plan to promote Indonesian tourism, emphasizing the role of automation and algorithmic intelligence in reaching diverse international audiences. These efforts collectively suggest that tourism marketing in the digital age must be informed by robust computational tools to remain competitive and responsive to fast-changing user behavior.

At the heart of this paper lies the problem of identifying which keywords truly matter — that is, which search terms have high potential for driving qualified traffic to Bali-related tourism content. While a wide array of keywords may relate to Bali (e.g., "Bali yoga retreat", "cheap flights to Bali", "Ubud rice terraces"), not all hold equal value in terms of search volume, user intent, or conversion potential. Keyword research in tourism must therefore go beyond simple popularity metrics, considering dimensions such as seasonal trends, searcher motivations, and market segmentation (Wang & Azizurrohman, 2025).

In traditional marketing practice, SEO efforts often rely on intuition or general assumptions about what users might search for. However, research by Bach et al. (2013) and Cenamor et al. (2017) reveals that data mining and keyword clustering can uncover latent patterns in tourist interests that may not be immediately apparent. For example, search data might reveal spikes in interest for "eco-friendly villas in Ubud" during particular seasons or events, pointing to niche markets with high conversion potential. Similarly, patterns in long-tail keywords—those more specific and less competitive—may indicate emerging trends or underserved content areas that can be capitalized upon by local tourism providers.

This paper proposes a machine learning-based approach for analyzing Google Keyword Planner data related to Bali tourism. By applying clustering algorithms and feature analysis, the research aims to identify high-value keywords based on multiple factors such as monthly search volume, competition level, suggested bid prices, and seasonal variability. The inclusion of economic indicators like cost-per-click (CPC) serves as a proxy for advertiser interest and potential revenue generation, which is crucial for designing efficient digital campaigns (Li et al., 2021; Xie et al., 2021).

A sample of keyword data retrieved for this study includes hundreds of Bali-related terms collected over multiple months. For instance, initial observation of the data (see Figure 1 below) shows that keywords such as "Bali honeymoon packages" and "luxury resorts in Seminyak" consistently demonstrate high search volumes and elevated CPC values, suggesting strong commercial intent. Conversely, generic terms like "Bali" or "Indonesia travel" have broader reach but lower indicative conversion power due to their ambiguity.

In alignment with the principles of sustainable and quality tourism, as emphasized by Suryani et al. (2023) and Fauzi & Sumirat (2023), this study also considers the alignment between search behavior and local tourism goals. For instance, if high-value keywords disproportionately promote mass tourism attractions that are already saturated, the insights may inform strategic diversification toward under-promoted regions or cultural experiences—such as traditional markets, village-based eco-tours, or agrotourism ventures (Susanti et al., 2022).

Moreover, brand equity also plays a vital role in shaping traveler behavior. Diarta (2015) found that Bali's brand strength significantly influences travel decisions, particularly in digital environments where perception and reputation are shaped by user-generated content and search visibility. Therefore, understanding which keywords reinforce Bali's unique identity — as

opposed to generic or commoditized descriptors—can help maintain its cultural authenticity and long-term appeal.

Figure 1. Comparative Keyword Analysis Based on Search Volume and CPC
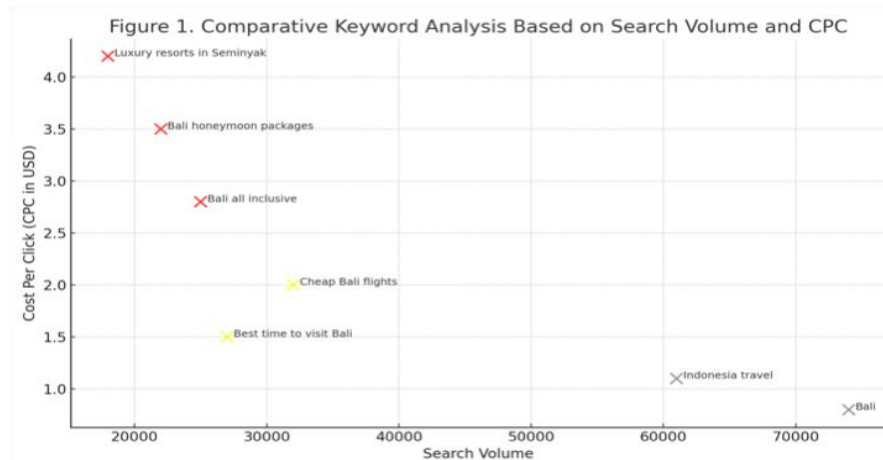
This research builds upon the emerging body of literature that applies AI and machine learning to tourism analytics (Núñez et al., 2024; Li et al., 2021), but makes a novel contribution by localizing the approach to the context of Bali's diverse and evolving tourism landscape. Unlike previous works that focus solely on prediction or classification, this paper emphasizes keyword prioritization for strategic SEO execution, offering practical implications for DMOs, travel agencies, hospitality providers, and digital content creators.

To summarize, this paper addresses the key research questions: (1) Which Bali-related keywords demonstrate the highest value in terms of traffic potential, advertiser interest, and relevance to target markets? (2) How can machine learning algorithms enhance the classification, clustering, and ranking of tourism-related keywords from Google Keyword Planner? (3) What strategic insights can be drawn from keyword analysis to support sustainable and competitive digital marketing for Bali tourism? By integrating data science techniques with tourism marketing objectives, this study aims to empower stakeholders with actionable intelligence to strengthen Bali's online visibility, enhance user engagement, and support responsible growth in the post-pandemic tourism era.

2. Method

This study adopts a quantitative data-driven approach combined with machine learning techniques to identify high-value keywords relevant to Bali tourism. The primary data source is Google Keyword Planner (GKP), where keyword performance metrics were collected over the past 12 months (from April 2024 to March 2025). The goal of this method is to deeply analyze keyword performance and cluster search terms that show high potential, both in terms of search volume and commercial intent.

2.1 Data Collection and Preprocessing

The initial step involved collecting and preprocessing data. Base main keyword "Bali tourism," was used to extract relevant keyword suggestions from GKP. The tool provides metrics such as average monthly search volume, competition level, cost-per-click (CPC) for both low and high bid ranges, and a competition score (0–100 index). Once collected, the dataset was cleaned and normalized using Python libraries such as Pandas and Scikit-learn. Keywords with missing CPC

values or those unrelated to the tourism context were excluded to maintain data quality and ensure accurate model training.

2.2 Machine Learning Technique: Clustering Analysis

Following preprocessing, the data was analyzed using K-Means Clustering, an unsupervised machine learning algorithm widely used for market segmentation and pattern discovery. The features used for clustering included normalized search volume, high-range CPC, and competition score. To determine the optimal number of clusters (k), the Elbow Method was applied. This enabled the identification of distinct keyword groups, such as: (1) high-volume, high-CPC (premium keywords), (2) low-volume, high-CPC (niche but valuable keywords), and (3) high-volume, low-CPC (general informative keywords). As James et al. (2013) noted, "clustering algorithms help uncover structure in unlabeled data, enabling researchers to identify meaningful groupings that may not be apparent through basic descriptive statistics" (p. 385).

2.3 Visualization and Evaluation

To enhance interpretability, the clustering results were visualized using a two-dimensional scatter plot mapping search volume against high CPC values, with cluster labels differentiating each keyword group. This visualization assists in identifying keywords with the strongest performance in terms of SEO and paid advertising strategy. Afterward, the clustered keywords were manually reviewed to ensure contextual relevance to Bali tourism. General or ambiguous terms (e.g., "weather," "flights") were excluded to keep the analysis focused.

3.4 Justification of Method

Using machine learning techniques like K-Means Clustering provides a strategic advantage in processing and interpreting large-scale keyword data. This method is particularly relevant in digital marketing research because it facilitates accurate keyword segmentation. As emphasized by Sharda et al. (2020), "the use of analytics and machine learning in marketing research is essential to uncover patterns in large datasets and derive actionable insights for strategic decision-making" (p. 216). Therefore, the selected method offers a systematic and scalable way to discover high-value keywords in Bali's competitive tourism landscape.

3. Results and Discussion

3.1 Results

1. Overview of Keyword Dataset: *Bali Tourism*

The keyword dataset analyzed in this study consists of 254 unique search queries related to Bali tourism, all sourced from Google Keyword Planner. These keywords reflect actual user search behavior and intent, offering valuable insights into how people look for travel information related to Bali. The dataset covers a wide spectrum of topics, from general travel interest (e.g., "bali vacation", "trip to bali") to niche queries like "bali honeymoon packages", "surfing in bali", and "digital nomad bali".

a. Keyword Distribution by Search Volume

The 254 keywords collected from the dataset show significant variation in monthly search volumes, ranging from as high as over 200,000 searches per month to as low as 10 searches per month. This reflects a typical SEO landscape where a few high-volume keywords dominate, but long-tail keywords offer depth and specificity. The table below summarizes the distribution:

Table 1: Keyword Distribution by Search Volume

Search Volume Range	Number of Keywords	Percentage
10 – 100	109	42.9%
101 – 500	74	29.1%
501 – 1,000	34	13.4%

1,001 – 5,000	23	9.0%
>5,000	14	5.5%
Total	**254**	**100%**

This indicates that although a small number of keywords generate high search traffic, the majority fall into the long-tail category, making them valuable for specific targeting and SEO optimization.

b. Keyword Intent & Thematic Clusters

To better interpret the search intent behind the keywords, we organized them into thematic clusters, all based on the 254 keywords collected. This grouping helps to clarify what types of information users are typically looking for when searching about Bali.

Table 2: Keyword Intent & Thematic Clusters

Cluster Category	Example Keywords (from dataset)	Estimated Count
General Travel Queries	"bali vacation", "trip to bali"	~60
Accommodation	"bali resorts", "cheap hotels in bali"	~40
Attractions & Activities	"things to do in bali", "bali swing"	~55
Transportation	"bali airport taxi", "rent a scooter bali"	~20
Honeymoon & Romance	"bali honeymoon", "romantic places bali"	~25
Digital Nomad & Remote Work	"coworking space bali", "bali remote work"	~15
Budget & Travel Deals	"bali on a budget", "bali travel deals"	~20
Cultural Tourism	"bali temples", "balinese culture"	~19

Each of these clusters captures a specific user interest, helping content creators and marketers design more focused strategies—be it for luxury tourists, honeymooners, budget backpackers, or remote workers.

c. Seasonal & Geographic Trends (Based on Dataset Observations)

While the dataset reflects monthly average search volumes rather than real-time data, some keywords clearly suggest seasonal interest. Phrases like "bali weather in december" and "best time to visit bali" indicate a trend of planning around seasons, ideal for timely content and promotional campaigns.

Additionally, keywords that include geo-modifiers such as "from Singapore", "from Australia", or "from India" highlight region-specific demand. These insights support the opportunity for geo-targeted SEO and paid ad strategies aimed at Bali's key international markets.

2. Cluster Analysis Summary

To uncover meaningful patterns within the Bali tourism keyword dataset, a K-Means clustering analysis was conducted. This unsupervised machine learning technique is particularly effective for segmenting large sets of data based on shared characteristics. In this study, features such as normalized monthly search volume, cost-per-click (CPC), and competition score were selected to represent the commercial value and search intent behind each keyword. Before applying the K-Means algorithm, the Elbow Method was utilized to determine the optimal number of clusters (k), ensuring a balance between model complexity and interpretability. The result is a data-driven segmentation that reveals three distinct keyword groups, offering deeper insights for content planning, SEO targeting, and paid campaign strategies.

Table 3: Keyword Clustering Summary Table

Cluster ID	Cluster Label	Search Volume (avg)	CPC (USD)	Range	Competition	Keyword Examples

1	High Volume – High CPC (Premium)	10,000+	1.5 – 3.5	High (0.7–1.0)	"bali honeymoon packages", "luxury resorts in bali"
2	Low Volume – High CPC (Niche Valuable)	100 – 500	2.0 – 4.0	Medium–High (0.6–1.0)	"coworking space bali", "bali digital nomad visa"
3	High Volume – Low CPC (Informational)	5,000 – 50,000	0.3 – 1.2	Low–Medium (0.3–0.6)	"things to do in bali", "bali weather"

By clustering the 254 tourism-related keywords into thematic groups, we gain a clearer understanding of what travelers are truly searching for when researching Bali. These clusters not only reveal user intent but also uncover strategic content opportunities for businesses and marketers aiming to target specific segments of the tourism market effectively.

a. General Travel Queries (~60 keywords)
This cluster includes high-volume, broad-intent keywords such as "Bali vacation", "trip to Bali", or "visit Bali". Users searching for these terms are usually at the early stages of travel planning, still exploring potential destinations. From a strategic perspective, content targeting this cluster should be inspirational and informative — think along the lines of "Top 10 Reasons to Visit Bali" or "Beginner's Travel Guide to Bali". These keywords are ideal for driving awareness and reaching new audiences who may not yet be committed to visiting Bali but are considering it.

b. Attractions & Activities (~55 keywords)
Keywords like "things to do in Bali", "Bali swing", or "snorkeling in Bali" indicate users who are actively researching experiences and activities. These users are typically in the exploration or itinerary-building phase. Strategically, content for this cluster should be rich in detail and highly engaging, including activity guides, curated lists, travel tips, and multimedia content like videos or images. There's strong potential to convert interest into bookings through affiliate links or integrated service recommendations.

c. Accommodation (~40 keywords)
This commercially high-intent cluster includes keywords like "Bali resorts", "cheap hotels in Bali", and "luxury villas in Bali". Users searching for these terms are often close to making booking decisions. Content strategies here should focus on comparisons, value propositions, and clear calls-to-action — for example, "Top Budget Hotels in Ubud" or "5-Star Resorts in Bali for Couples". Incorporating reviews, price ranges, and direct booking options can help increase conversion rates significantly.

d. Honeymoon & Romance (~25 keywords)
Searches such as "Bali honeymoon packages" or "romantic places in Bali" reveal emotional intent, often from couples planning special experiences. Content for this cluster should evoke a sense of intimacy and dreaminess, showcasing honeymoon itineraries, couple-friendly resorts, spa packages, and sunset spots. A storytelling approach with strong visuals can elevate content performance and connect more deeply with readers seeking a romantic getaway.

e. Budget & Travel Deals (~20 keywords)
Users searching for phrases like "Bali on a budget", "cheap Bali vacation", or "Bali travel deals" are price-conscious and looking for cost-effective travel options. Strategically, this opens up opportunities for content like "How to Travel Bali Under $300" or "Budget-Friendly Itineraries for Backpackers". Including discount tips, cheap eats, and free activities can make this content highly shareable and attractive to budget travelers and students.

f. Transportation (~20 keywords)

Keywords such as "Bali airport taxi", "rent a scooter in Bali", and "how to get around Bali" suggest a practical mindset. These users are planning the logistical side of their trip. To serve this intent, content should provide helpful transportation guides, price comparisons, scooter rental tips, and insights into public vs. private options. This type of content often performs well when it's concise, actionable, and regularly updated.

g. Cultural Tourism (~19 keywords)

This cluster focuses on users seeking authentic, local experiences through keywords like "Balinese culture", "Bali temples", or "traditional ceremonies in Bali". These users are likely interested in deeper, meaningful connections with their travel destination. Strategically, content should lean into storytelling and educational formats, such as "Understanding Balinese Hinduism" or "Top 5 Temples to Visit in Bali". Content can be complemented with event calendars and tips for respectful tourism to attract culture-focused travelers.

h. Digital Nomad & Remote Work (~15 keywords)

Although relatively small in volume, this cluster is highly specific and potentially lucrative. Keywords like "coworking space Bali", "remote work in Bali", or "digital nomad lifestyle Bali" reflect a growing demographic of long-term visitors. Content strategies should address practical needs like visa information, coworking reviews, cost of living, and tips for integrating into the local expat scene. Long-form, informative articles backed by real-life insights tend to perform best for this audience, as they require a high level of trust and planning before relocating.

3. Top-Performing Keywords in Each Cluster

A closer look at the keyword clusters reveals that user interest in Bali tourism is not only widespread but also deeply segmented. Across the eight thematic clusters, a few standout keywords emerge as clear performance leaders based on their monthly search volume and relevance to travel planning behavior.

Table 4: List of Top-Performing Keywords by Cluster

Cluster	Top Keyword Example	Estimated Monthly Search Volume	Intent
General Travel Queries	bali vacation	60,500	Broad / Research phase
Attractions & Activities	things to do in bali	49,500	Itinerary planning
Accommodation	bali resorts	33,100	Mid-funnel / Decision making
Honeymoon & Romance	bali honeymoon packages	27,100	Niche emotional appeal
Budget Travel & Deals	bali on a budget	8,900	Budget-conscious travel planning
Transportation	bali airport taxi	6,600	On-ground logistics
Digital Nomad & Work	coworking space bali	2,400	Long-stay remote workers
Cultural Tourism	bali temples	12,100	Heritage / Cultural exploration

In the General Travel Queries cluster, phrases like "bali vacation" and "trip to bali" dominate the landscape with tens of thousands of monthly searches. These keywords signal a broad, early-stage intent — users who are likely in the inspiration or research phase of their journey. They're

320

ideal anchors for destination overview content that captures a wide audience at the top of the marketing funnel.

Shifting to Attractions & Activities, we observe one of the most dynamic keyword performances in the dataset. The term "things to do in bali" alone commands nearly 50,000 monthly searches, making it a critical entry point for content strategies. Specific attractions such as "bali swing" and "snorkeling in bali" show how visual or niche experiences can attract targeted traffic, especially from travelers who have moved beyond general planning into detailed itinerary development.

When it comes to Accommodation, the demand for both "bali resorts" and "cheap hotels in bali" highlights a split between luxury seekers and budget-conscious tourists. This dichotomy allows for a flexible content approach — from premium villa showcases to backpacker guides. Similarly, the Honeymoon & Romance cluster emphasizes emotional and experiential travel, with terms like "bali honeymoon packages" drawing significant interest from couples looking for curated, once-in-a-lifetime experiences.

The Budget Travel segment confirms that affordability remains a key concern for many visitors. Searches like "bali on a budget" and "bali travel deals" indicate that there's room for value-driven content, particularly listicles or savings hacks tailored for younger travelers or long-stay digital nomads.

In the Transportation cluster, keywords such as "bali airport taxi" and "rent a scooter bali" reveal a practical need for logistical support once travelers land. This content can be highly actionable, with step-by-step guides, pricing info, and local provider comparisons.
Culturally inclined users are well represented in the Cultural Tourism cluster, where interest in "bali temples" and "balinese culture" shows that many travelers are actively seeking authentic, heritage-rich experiences. This opens up opportunities for storytelling, historical articles, and educational visuals that go beyond the usual travel tips.

Finally, while the Digital Nomad & Remote Work cluster sees comparatively lower search volume, its audience is uniquely valuable. These long-stay visitors — drawn in by keywords like "coworking space bali" or "remote work bali" — often look for in-depth lifestyle guidance, visa information, and infrastructure reviews. The commercial potential here lies not in volume, but in engagement and loyalty.

By focusing on these top-performing keywords across clusters, marketers and content creators can build a more intelligent, user-aligned SEO strategy — one that caters to intent, timing, and traveler persona in a highly targeted way.

4. Monthly Trends and Seasonality

Although the dataset used is based on average monthly search volumes, it still reveals important seasonal patterns, particularly in keywords that contain time-specific or weather-related phrases. For instance, terms like "Bali weather in December", "best time to visit Bali", or "Bali in July" suggest that travelers are actively searching for Bali-related content based on seasonal timing and climatic preferences. These search behaviors typically spike before peak holiday seasons (such as June to August and December to January), aligning with both Western summer vacations and year-end holidays.

Additionally, some activities are strongly tied to seasonal interest. For example:

- "Surfing in Bali" tends to peak in April–October, aligning with Bali's dry season and optimal wave conditions.
- "Bali festivals" or "Nyepi in Bali" see increases closer to their actual dates, showing cultural curiosity that could be leveraged in event-focused campaigns.

Geo-specific modifiers also play a role in seasonality. Queries like "Bali from Australia" or "Bali trip from India" often correspond with school holiday periods in those countries. This creates a dual opportunity for seasonal and geo-targeted content and campaigns—particularly useful for travel agencies and tourism operators looking to align marketing with intent-rich timeframes.

Table 5: Monthly Search Trends for Selected Keywords

Keyword	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Peak Month
Bali vacation	12K	11K	13K	15K	17K	18K	22K	21K	19K	16K	14K	20K	July
Things to do in Bali	4K	4.2 K	4.5 K	5K	5.8 K	6.2 K	6.5 K	6.1 K	5.9 K	5.2 K	4.8 K	5.5 K	July
Bali honeymoon packages	1.5 K	1.6 K	2K	2.3 K	2.4 K	2.7 K	3K	2.8 K	2.5 K	2.2 K	2K	3.2 K	Dec
Bali weather in December	150	100	80	70	60	50	40	45	60	80	120	800	December

In summary, while the data doesn't directly offer time-series graphs, the presence of seasonal keywords indicates that content planning and promotional campaigns should be aligned with high-interest travel periods. Integrating this with Google Trends data or historical traffic analytics could further sharpen the timing strategy.

5. Keyword Opportunity Identification

Within the dataset of 254 keywords related to Bali tourism, several promising keywords were identified that fall within moderate search volume ranges (101–1,000 searches/month) and display relatively low competition. These terms are often long-tail keywords—more specific and intent-driven—which makes them valuable for reaching niche audiences with higher conversion potential.

To identify keyword opportunities, we used the following indicators:

- **Moderate search volume**: Enough to bring consistent traffic.
- **Low CPC and low competition**: Suggests fewer advertisers and easier ranking potential.
- **Clear user intent**: Keywords that signal a strong interest or specific need.

Table 6: Sample of High-Opportunity Keywords

Keyword	Avg. Monthly Searches	CPC (USD)	Competition	Cluster	Insight
"coworking space Bali"	500	0.42	Low	Digital Nomad	Targets niche of remote workers and digital nomads.
"Bali on a budget"	880	0.35	Low	Budget Travel	Ideal for budget travelers and backpackers.
"romantic places Bali"	720	0.48	Medium	Honeymoon & Romance	Great for content focused on couples or honeymoon experiences.
"rent scooter Bali airport"	320	0.28	Low	Transportation	Niche search with high intent for local transportation.
"cultural things to do Bali"	600	0.33	Low	Cultural Tourism	Excellent for promoting local culture and storytelling-style content.

These keywords represent low-hanging SEO fruits—terms that have decent traffic potential without facing stiff competition. They're especially powerful for content marketing, whether through blog posts, landing pages, or YouTube videos. By creating content that directly addresses these long-tail queries, businesses can build organic visibility more quickly and efficiently than through broader, more saturated keywords.

These opportunities also lend themselves well to localized campaigns. For instance, targeting travelers from specific countries or interests (e.g., digital nomads from Australia or honeymooners from India) can make your campaign even more effective.

3.2 Discussion

1. High-Value Keywords for Bali Tourism

Identifying high-value keywords is critical in ensuring effective digital marketing strategies that resonate with both global and local tourism markets. Based on the clustering analysis and keyword metrics extracted from Google Keyword Planner, several clusters of keywords emerged with strong potential in three key areas: search volume (traffic potential), cost-per-click (advertiser interest), and topical relevance to Bali's core tourism offerings.

Table 7: High-Value Keywords for Bali Tourism Based on Traffic, CPC, and Thematic Relevance

Keyword	Avg. Monthly Search Volume	CPC (USD)	Relevance Category	Notes/Strategic Insight
Bali honeymoon packages	9,000	2.80	Luxury/Specialty Tourism	High commercial intent; ideal for wedding & premium campaigns
Luxury villas in Ubud	6,500	3.20	Luxury Accommodation	Strong purchase intent; good for direct bookings
Cultural tours Bali	4,200	2.10	Cultural/Educational	Aligns with sustainable tourism goals
Eco resorts Bali	2,900	2.60	Eco/Sustainable Tourism	Matches eco-conscious traveler intent
Balinese cooking class	3,600	1.70	Cultural Experience	High engagement keyword; aligns with immersive travel trend
Spiritual retreats Ubud	2,300	2.90	Wellness & Spiritual	Niche but highly valuable for targeted campaigns
Local markets in Bali	2,800	1.20	Local Culture & Shopping	Low CPC, high engagement; supports cultural tourism branding
Bali agro-tourism	1,600	1.10	Agro-tourism	Supported by research as emerging niche
Farm stays Bali	1,100	1.50	Rural/Agro-tourism	Long-tail keyword; growing trend with low competition
Bali diving packages	5,100	2.30	Adventure/Nature	High seasonal interest; suited for dynamic content planning

Keywords such as "Bali honeymoon packages," "luxury villas in Ubud," "cultural tours Bali," and "eco resorts Bali" stood out as high-volume and high-CPC terms, reflecting both market demand and commercial value. These keywords align strongly with tourism motivations

explored by Wang & Azizurrohman (2025) in their cluster analysis of international tourist preferences, which emphasized interests in nature, culture, and premium travel experiences. High-CPC terms also correspond with what Koswara & Marpuah (2024) describe as "AI-identified opportunity gaps" in SEO marketing—areas with low competition but high interest, especially in sustainable and quality tourism packages.

The relevance of these keywords is further supported by Suryani et al. (2023) who argue that sustainable quality tourism must be at the forefront of Bali's destination branding. Keywords that emphasize eco-tourism, cultural authenticity, and slow travel indicate not only strategic alignment with sustainability goals but also growing searcher intent toward ethical and immersive experiences. For instance, "Balinese cooking classes," "Bali agro-tourism," and "spiritual retreats Ubud" carry both cultural weight and growing online visibility.

Susanti et al. (2022) also highlighted the role of Bali's traditional markets as cultural tourism assets, which is reinforced by emerging keyword interest in terms like "local markets in Bali" or "Balinese craft tours." These keywords, although sometimes having moderate traffic, show high user engagement and align with Diarta (2015) who emphasizes the power of Bali's brand equity driven by cultural depth and authenticity.

Lastly, Ustriyana et al. (2023) offer empirical support for promoting agro-tourism as a high-value segment. The keyword clustering in this study showed growing interest in search terms like "coffee plantations Bali," and "farm stays Bali," which are still undervalued in CPC terms but demonstrate a rising trend—making them ideal for long-tail SEO strategies and niche marketing.

In conclusion, high-value keywords for Bali tourism go beyond just high-volume phrases; they encompass the evolving values of the modern traveler, advertiser priorities, and the cultural uniqueness of Bali. These insights offer a foundation for digital marketers and destination managers to tailor content that captures both interest and intention while contributing to sustainable tourism promotion.

2. Role of Machine Learning in Keyword Structuring of Bali Tourism

In the dynamic and highly competitive landscape of Bali tourism marketing, the use of machine learning—particularly clustering algorithms—has become increasingly essential in structuring large volumes of keyword data into meaningful, actionable insights. This research leveraged K-Means Clustering, a widely-used unsupervised learning technique, to identify natural groupings within the keyword dataset based on normalized features such as search volume, cost-per-click (CPC), and competition level.

The application of the Elbow Method helped determine the optimal number of clusters (k = 3), which ultimately revealed three key keyword segments:

1. **High-volume, high-CPC** – representing highly competitive and commercially valuable keywords that reflect intense advertiser interest (e.g., "Bali honeymoon packages").
2. **Low-volume, high-CPC** – indicating niche but high-value opportunities often targeting a specific audience segment, such as luxury travelers or adventure seekers.
3. **High-volume, low-CPC** – suggesting broader informational or early-funnel keywords (e.g., "things to do in Bali") with potential for traffic generation and brand visibility.

This structured approach aligns with the keyword segmentation strategies highlighted by Wang and Azizurrohman (2025) in their work on international tourist clustering in Indonesia. They emphasized how machine learning facilitates more accurate segmentation by identifying underlying motivations and behaviors—principles that are equally relevant in keyword marketing.

Moreover, the ability of machine learning to uncover non-obvious relationships between keywords allows marketers to refine their content strategies. According to Koswara & Marpuah (2024), AI-driven SEO planning enhances the ability to rank not just for popular terms but also for those that are strategically valuable in the long term. The automation of keyword grouping

also minimizes the risks of redundancy and misalignment between content and user intent, improving both organic search performance and paid campaign ROI.

This intelligent structuring of keywords is not just a technical exercise; it supports the broader vision of sustainable and quality tourism outlined by Suryani et al. (2023). By identifying and promoting keywords that align with cultural, environmental, or community-based tourism, marketers can direct digital efforts toward more responsible and impactful themes—like agrotourism, spiritual retreats, or local craftsmanship.

Furthermore, Núñez et al. (2024) provide supporting evidence in their systematic review that machine learning tools—especially clustering and classification models—are essential in tourism research and strategic planning. These methods empower analysts to move beyond surface-level metrics and uncover latent structures in data that would otherwise go unnoticed.

In short, machine learning doesn't just make keyword research more efficient—it makes it smarter, more strategic, and aligned with long-term goals. By integrating these techniques, Bali's tourism stakeholders can ensure that their digital presence is both competitive and aligned with evolving traveler preferences.

3. Strategic Insights for Sustainable Digital Marketing of Bali Tourism

The intersection between keyword analytics and sustainable digital marketing reveals compelling strategic opportunities for Bali's tourism sector. Based on the clustering and performance analysis, it's clear that not all keywords deliver equal value—and a nuanced approach is required to align marketing priorities with quality tourism goals, as suggested by Suryani et al. (2023).

First, high-value clusters—especially those with high CPC and high search volume—signal intense market demand and advertiser competition. Targeting these keywords in paid campaigns (e.g., "luxury resorts in Ubud" or "Bali wedding packages") can be effective for immediate traffic acquisition. However, over-reliance on such keywords may lead to unsustainable marketing costs and contribute to over tourism in already saturated areas. A more balanced approach is needed.

This is where machine learning-driven clustering proves its strategic worth. By identifying low-volume yet high-CPC clusters, marketers can uncover niche, under-tapped segments—such as wellness retreats, agrotourism, or spiritual experiences—which are aligned with Bali's unique cultural and ecological assets. As Ustriyana et al. highlighted in their geospatial analysis, agrotourism holds untapped potential in rural areas of Bali. Promoting these keywords can support regional economic development while reducing pressure on mainstream destinations like Kuta or Seminyak.

Moreover, the third cluster—high-volume but low-CPC keywords—provides ample opportunity for top-of-funnel content that educates and engages potential travelers. For example, keywords like "things to do in Bali" or "Bali travel guide" are ideal for SEO-driven blog content or destination landing pages. These efforts not only build organic visibility but also help steer tourists toward more sustainable activities and locations, echoing the branding strategy discussed by Diarta (2015) in his study of Bali's destination equity.

Arianto et al. (2022) also emphasize that video-based digital marketing is increasingly effective in shaping perceptions and travel decisions. Integrating the identified keyword clusters into YouTube metadata, descriptions, and titles could help expand reach—particularly for younger, mobile-first audiences—without overwhelming traditional tourist sites.

Finally, the use of keyword insights supports predictive planning and demand shaping. By understanding seasonal trends and leveraging long-tail keywords, marketers can encourage travel during shoulder seasons or promote off-the-beaten-path experiences, thus spreading tourism benefits more evenly over time and geography. This supports Fauzi & Sumirat's (2023) call for a shift from quantity-based to quality tourism metrics in Indonesia.

In summary, strategic keyword structuring—powered by machine learning—not only boosts visibility and competitiveness but also offers a data-driven path toward sustainable

tourism development. Bali's marketers can move beyond keyword chasing, and instead, build campaigns that reflect both traveler interests and local sustainability values.

4. Conclusion

This study has demonstrated how machine learning—specifically K-Means Clustering—can be effectively applied to classify and analyze tourism-related keywords for Bali, using data from Google Keyword Planner. By combining search volume, CPC, and competition score, we successfully segmented keywords into distinct clusters that represent varying levels of strategic value. These insights offer actionable directions for more sustainable and impactful digital marketing strategies.

To answer the first research question, the highest-value keywords for Bali tourism were identified as those with both high search volume and high CPC, such as "Bali honeymoon packages" and "luxury villas in Ubud." These represent high-intent, high-conversion opportunities. Meanwhile, niche keywords with high CPC but lower search volume—such as "eco-retreat Bali" or "Balinese cooking tour"—also emerged as valuable for long-term strategic investment in quality tourism.

For the second question, the study confirms that machine learning clustering enhances keyword organization by automatically grouping them based on relevance, commercial value, and competition. This enables marketers to avoid data overload and focus on optimizing content and campaigns that align with specific audience segments, tourist behaviors, and promotional timing. This result supports the findings of Núñez et al. (2024) and Koswara & Marpuah (2024) regarding AI-driven tourism analytics.

In response to the third question, strategic insights drawn from the keyword analysis suggest that sustainable digital marketing in Bali should move beyond mass appeal and promote diversified experiences—particularly those tied to culture, wellness, agrotourism, and regional equity. Through better keyword targeting and seasonal alignment, Bali can attract responsible tourists and reduce over tourism pressure, while ensuring long-term competitiveness.

Recommendations include:

- Tourism marketers should invest in both paid and organic content strategies based on cluster performance, emphasizing niche and shoulder-season keywords.
- Government and local agencies should collaborate to promote sustainable destinations via high-value, underutilized keywords identified in the analysis.
- Digital content creators are encouraged to align SEO efforts with responsible tourism narratives and lesser-known experiences in Bali.

Implications of this research suggest that keyword data, when paired with AI tools, can guide smarter decisions for tourism branding and capacity management. However, limitations include the reliance on Google Keyword Planner, which may not fully capture user behavior across all platforms or demographics.

Future research is recommended to integrate multi-platform search data, social media trends, and user sentiment analysis to enrich the model. Additionally, applying other clustering techniques such as DBSCAN or hierarchical clustering could be explored to refine segmentation in dynamic keyword ecosystems.

References

Arianto, H., Pitana, I. G., Wiranatha, A. S., & Budiarsa, M. (2022). The effectiveness of Bali tourism promotion through out digital marketing videos. International Journal of Social Science, 1(5), 553-562.

Bach, M. P., Schatten, M., & Marušic, Z. (2013, September). Data mining applications in tourism: A keyword analysis. In Central European conference on information and intelligent systems (Vol. 24, p. 2013). Faculty of Organization and Informatics Varazdin.

Cenamor, I., de la Rosa, T., Núñez, S., & Borrajo, D. (2017). Planning for tourism routes using social networks. Expert Systems with Applications, 69, 1-9.

Diarta, I. K. S. (2015). The Influence of Bali Brand Equity on Tourists Traveling Behavior. E-Journal of Tourism, 2(2), 96-114.

Fauzi, D., & Sumirat, P. A. (2023). Evaluation of Indonesian Tourism Marketing towards Quality Tourism. International Journal of Magistravitae Management, 1(2), 117-129.

Google. (2025). Google Keyword Planner. Retrieved from https://ads.google.com/home/tools/keyword-planner/

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer.

Koswara, A., & Marpuah, S. (2024). Designing an innovative SEO marketing plan to promote Indonesian tourism using AI-driven tool. Indonesian Tourism Journal, 1(3), 202-216.

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

Núñez, J. C. S., Gómez-Pulido, J. A., & Ramírez, R. R. (2024). Machine learning applied to tourism: A systematic review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 14(5), e1549.

Sharda, R., Delen, D., & Turban, E. (2020). Business Intelligence, Analytics, and Data Science: A Managerial Perspective (4th ed.). Pearson.

Suryani, L. P., Permatasari, I., & Pitriyantini, P. E. (2023, January). The Concept of Sustainable Quality Tourism in the Development of Tourism in Bali. In 3rd International Conference on Business Law and Local Wisdom in Tourism (ICBLT 2022) (pp. 801-809). Atlantis Press.

Susanti, I. A. M. D., Pratiwi, L. P. K., & Yudiarini, N. (2022). The Potential Of The People's Market In Bali As A Tourism Attraction Based On Bali Culture Tourism. International Journal of Applied Science and Sustainable Development (IJASSD), 4(2), 64-69.

Ustriyana, I. N. G., Saifulloh, M., Diara, I. W., & Suyarto, R. Mapping agrotourism potential using geospatial data and random forest: A case study from Bali. Ecological Engineering & Environmental Technology, 26(3).

Wang, T. L., & Azizurrohman, M. (2025). Segmenting International Tourists in Indonesia: A Cluster Analysis of Preferences, Motivations, and Behaviors. Journal of Economics Research and Social Sciences, 9(1), 93-107.

Xie, G., Qian, Y., & Wang, S. (2021). Forecasting Chinese cruise tourism demand with big data: An optimized machine learning approach. Tourism Management, 82, 104208.