Using K-Means Clustering to Enhance Digital Marketing with Flight Ticket Search Patterns

Husni Teja Sukmana[1,*], Lee Kyung Oh[2]

[1]Syarif Hidayatullah State Islamic University Jakarta

[2] Sun Moon University Asan, Republic of Korea

ABSTRACT

This study explores the application of K-Means clustering to enhance digital marketing strategies by analyzing flight ticket search patterns. Utilizing a dataset containing 4,000 search engine results related to flights to Hong Kong, the research identifies five distinct user clusters based on search terms, titles, snippets, and other relevant features. The dataset's key features include search terms, ranks, titles, snippets, display links, and direct links, providing a comprehensive view of user interactions and preferences. The cluster analysis reveals significant variations in user intent and preferences across the identified segments. For instance, Cluster 1 is characterized by users searching for "cheap flights" and "discount tickets," indicating a price-sensitive segment. In contrast, Cluster 2 users prefer "premium flights" and "business class," highlighting an interest in luxury travel options. The study also examines the behavioral patterns within each cluster, such as Cluster 3 users who search for flights well in advance and prioritize flexible booking options. The findings underscore the effectiveness of K-Means clustering in enhancing digital marketing strategies. By leveraging the insights from the clustering analysis, marketers can design highly targeted advertising campaigns and personalized offers. For example, budget airlines can target Cluster 1 with promotions and discounts, while premium airlines can focus on Cluster 2 with exclusive service highlights. This targeted approach is expected to improve user engagement and conversion rates significantly. The study also highlights the advantages of behavior-based segmentation over traditional demographic methods, offering a more accurate representation of user preferences and intentions. The identified clusters provide a framework for understanding different user groups, enabling more efficient resource allocation and campaign design. Future research should explore the integration of additional data sources, such as social media interactions and user reviews, to enhance clustering accuracy. Additionally, advanced clustering techniques like hierarchical clustering and Gaussian Mixture Models could be investigated to provide further insights. The ongoing refinement and enhancement of segmentation processes are crucial for maintaining effective and impactful digital marketing strategies in the dynamic travel industry. Key results include the identification of five user clusters, the importance of personalized marketing strategies, and the potential for improved engagement and conversion rates through targeted advertising and offers.

Keywords K-Means Clustering, Digital Marketing, Flight Ticket Search Patterns, User Segmentation, Targeted Advertising, Personalized Offers, User Behavior Analysis, Travel Industry, Search Engine Data, Marketing Strategies

INTRODUCTION

In today's interconnected world, digital marketing has become a cornerstone of the travel industry. The advent of the internet and the proliferation of digital devices have revolutionized how travel services are marketed, consumed, and perceived. Travelers now have unprecedented access to information and

Submitted 19 September 2024
Accepted 23 October 2024
Published 1 December

Corresponding author
Husni Teja Sukmana,
husniteja@uinjkt.ac.id

Additional Information and Declarations can be found on page 302

DOI: 10.47738/jdmdc.v1i3.22

How to cite this article: H. T. Sukmana, L. K. Oh, "Using K-Means Clustering to Enhance Digital Marketing with Flight Ticket Search Patterns," *J. Digit. Mark. Digit. Curr.*, vol. 1, no. 3, pp. 286-304, 2024.

options, allowing them to make more informed decisions about their travel plans. Digital marketing enables travel companies to reach a global audience with targeted and personalized messages, enhancing customer engagement and satisfaction. This shift from traditional to digital marketing has not only increased the efficiency of reaching potential customers but also significantly reduced the costs associated with marketing campaigns.

Digital marketing has significantly transformed the travel industry, reshaping how destinations engage with travelers and driving economic growth [1]. The transition from traditional to digital marketing methods, particularly through search engines and social media, has been substantial for travel agents, enterprises, and destinations [2]. The integration of digital innovations has become essential in modern tourism marketing efforts, enabling destinations to utilize various tools and platforms to engage audiences and inspire travel [3]. Techniques such as SEO, content marketing, and social media advertising have empowered tourism businesses to enhance their online presence and effectively connect with potential travelers [4].

Social media plays a crucial role in enhancing destination image and promoting tourism, especially in challenging situations like political conflicts [5]. The use of digital marketing platforms such as websites, social media, and online advertising has facilitated the marketing activities of tourism destinations, making them easily accessible to a broader audience [6]. Additionally, the adoption of digital marketing by Islamic tourism organizations has been studied, emphasizing the importance of understanding factors affecting digital marketing adoption and implementation for these organizations [7].

In the digital transformation era, digital tourism marketing is vital in facilitating travel processes for tourists and benefiting the entire tourism industry through improved communication and transactions [8]. The use of digital marketing tools in the tourism sector, such as websites and social media, has not only changed industry paradigms but also transformed communication, transactions, and lifestyle aspects [9]. Overall, the integration of digital technologies in tourism marketing has reshaped how companies communicate with their target audiences and promote their products and services [10].

Moreover, digital marketing strategies in the travel industry leverage various channels such as social media, email marketing, search engine optimization (SEO), and pay-per-click (PPC) advertising. These strategies are designed to attract, engage, and retain customers by providing relevant content and offers tailored to their preferences and behaviors. The ability to track and analyze customer interactions through digital channels has given travel companies invaluable insights into consumer behavior, allowing for more effective marketing and improved customer experiences. As a result, digital marketing has become indispensable for the travel industry, driving growth and innovation in a highly competitive market.

With the rise of digital marketing, the importance of search engine data has grown exponentially. Search engines are often the starting point for consumers planning their travel, making search engine data a goldmine for marketers looking to understand consumer intent and preferences. By analyzing search queries, travel companies can gain insights into what potential travelers are looking for, allowing them to tailor their marketing strategies to meet these needs more effectively. This data-driven approach not only enhances the precision of marketing campaigns but also increases the likelihood of converting searchers into customers.

Search engine data plays a crucial role in optimizing various aspects of digital marketing strategies. SEO is a fundamental technique that involves a combination of methods to ensure web pages are indexed, develop relevant content, and encourage click-throughs using specific keywords [11]. The use of Data Sciences has significantly increased in the digital marketing environment, aiding in decision-making and extracting actionable insights from large datasets [1]. SEO is essential for improving website visibility in search engine results, which is vital for enhancing brand image, attracting quality traffic, and safeguarding reputation [12].

SEO is a common strategy recognized globally for enhancing online content visibility and search engine rankings [13]. It focuses on optimizing digital assets to achieve higher rankings on search engine results pages, ultimately driving organic traffic to websites [14]. Additionally, SEO is crucial in promoting and marketing tourism destinations, emphasizing the importance of optimizing web content for better search engine visibility [15].

Search engine data provides a wealth of information that can be used to optimize various aspects of digital marketing. For instance, keyword analysis helps identify the terms and phrases most commonly used by potential travelers, informing SEO strategies to improve visibility in search engine results pages (SERPs). Additionally, analyzing search trends over time can reveal seasonal patterns and emerging travel destinations, enabling companies to adjust their marketing efforts accordingly. The ability to track and measure the performance of keywords and campaigns in real-time further allows for continuous optimization and refinement of marketing strategies, ensuring they remain relevant and effective in an ever-changing digital landscape.

In the travel industry, the relevance of analyzing search engine metadata for flight tickets cannot be overstated. Search engine metadata, which includes search terms, click-through rates, and user interaction metrics, provides a rich source of information about consumer behavior and preferences. By examining this data, travel companies can gain deep insights into the demand for specific flight routes, preferred travel times, and pricing sensitivity. This understanding allows companies to tailor their offerings to better meet customer needs, thereby enhancing their competitive advantage.

Furthermore, search engine metadata helps in identifying emerging trends and shifts in consumer behavior. For instance, an increase in search queries for flights to a particular destination might indicate a surge in interest due to events, seasonal changes, or promotions. By analyzing this data, airlines and travel agencies can adjust their marketing strategies, allocate resources more efficiently, and optimize inventory management to capitalize on these trends. Additionally, this analysis can aid in improving the user experience on booking platforms by making search results more relevant and personalized, ultimately leading to higher conversion rates and customer satisfaction.

To effectively leverage the insights derived from search engine metadata, advanced data analysis techniques are essential. One such method is K-Means clustering, a popular machine learning algorithm used to segment data into distinct groups based on their characteristics. In the context of flight ticket searches, K-Means clustering can be employed to group users with similar search behaviors and preferences. This segmentation allows travel companies to develop more targeted marketing strategies and personalized recommendations, thereby enhancing customer engagement and loyalty.

K-Means clustering works by partitioning the dataset into K clusters, where each cluster contains data points that are more similar to each other than to those in other clusters. This method is particularly useful for handling large datasets with complex patterns, as it can efficiently identify and group similar search queries. For example, clusters might reveal groups of users searching for budget flights, premium flights, or last-minute deals. By understanding these segments, travel companies can tailor their advertising and promotional efforts to address the specific needs and preferences of each group, resulting in more effective and impactful marketing campaigns.

The primary objective of this study is to explore how K-Means clustering can be utilized to enhance digital marketing strategies by analyzing patterns in flight ticket searches. In the competitive landscape of the travel industry, understanding user behavior is paramount to creating effective marketing campaigns. K-Means clustering, a robust unsupervised machine learning algorithm, provides a method to segment users based on their search patterns. By grouping users with similar behaviors, marketers can tailor their strategies to meet the specific needs of different user segments, thereby increasing engagement and conversion rates.

This paper aims to demonstrate the practical application of K-Means clustering in analyzing search engine metadata related to flight tickets. By identifying distinct clusters of search behaviors, we can uncover valuable insights into user preferences and trends. For instance, certain clusters might reveal users who consistently search for budget flights, while others may indicate a preference for premium travel options or last-minute deals. Understanding these segments allows marketers to design more personalized and targeted advertising campaigns, which can significantly improve the effectiveness of digital marketing efforts.

In addition to enhancing targeted marketing, K-Means clustering can also help optimize resource allocation and campaign management. By recognizing which user segments are more likely to convert, travel companies can allocate their marketing budgets more efficiently, focusing on high-potential customers. This strategic approach not only maximizes the return on investment (ROI) but also improves the overall customer experience by delivering relevant content and offers. Through this study, we aim to provide a comprehensive analysis of how K-Means clustering can be leveraged to gain actionable insights from search engine metadata, ultimately driving more informed and effective digital marketing strategies in the travel industry.

Literature Review

Digital Marketing in the Travel Industry

The travel industry has undergone a significant transformation over the past decade, largely driven by advancements in digital marketing. One of the most prominent trends is the increasing use of personalized marketing strategies.

With the rise of big data and machine learning, travel companies can now leverage vast amounts of user data to create highly targeted and personalized marketing campaigns. This shift has been facilitated by the proliferation of digital channels such as social media, email, and search engines, which offer platforms for reaching consumers in more engaging and relevant ways. Personalized marketing has shown to increase customer satisfaction and loyalty, as it addresses individual preferences and needs more effectively than traditional mass marketing approaches.

However, the adoption of digital marketing in the travel industry is not without its challenges. One of the main issues is the high level of competition. With numerous travel companies vying for the same audience, standing out requires innovative and compelling marketing strategies. Additionally, the rapid pace of technological change means that travel marketers must continually adapt to new tools and platforms. Privacy concerns and data protection regulations, such as the GDPR in Europe, also pose significant challenges. These regulations require companies to handle customer data responsibly and transparently, adding another layer of complexity to digital marketing efforts.

Data analytics plays a pivotal role in shaping digital marketing strategies in the travel industry. By analyzing customer data, travel companies can gain insights into consumer behavior, preferences, and trends. This information is crucial for creating effective marketing campaigns that resonate with target audiences. For instance, data analytics can reveal patterns in booking behaviors, such as preferred travel times, popular destinations, and price sensitivity. These insights allow marketers to tailor their offerings and promotions to align with consumer expectations, thereby increasing the likelihood of conversion.

Moreover, data analytics enables real-time decision-making, which is essential in the fast-paced travel industry. By continuously monitoring and analyzing data, companies can quickly adjust their marketing strategies in response to changing market conditions and consumer behaviors. This agility is particularly important in dealing with unexpected events, such as political unrest or natural disasters, which can significantly impact travel plans. Additionally, advanced analytics techniques, such as predictive modeling and sentiment analysis, help marketers anticipate future trends and customer needs, allowing them to stay ahead of the competition. Overall, data analytics not only enhances the effectiveness of digital marketing campaigns but also drives business growth by providing a deeper understanding of the market and the consumer.

Search Engine Metadata Analysis

The analysis of search engine metadata has become a crucial tool for gaining market insights across various industries. Previous studies have demonstrated that search data can effectively reflect consumer interest, predict market trends, and inform strategic decisions. For instance, a study highlighted the predictive power of Google search queries in forecasting economic indicators such as automobile sales, home sales, and travel trends [16]. Their research illustrated that the volume of specific search terms correlated strongly with real-world economic activities, providing a timely and cost-effective means of market analysis.

Further research has explored the application of search data in more specialized contexts. For example, [17] investigated the use of search engine data to understand consumer decision-making processes in online travel planning. Their findings indicated that search behaviors, such as the frequency and type of queries, could reveal significant information about consumer preferences and the stages of the decision-making process. Similarly, [18] examined the relationship between search engine data and brand metrics, demonstrating that search query volumes could serve as reliable proxies for brand interest and market share. These studies collectively underscore the value of search engine metadata as a rich source of real-time market intelligence.

The benefits of analyzing search engine metadata are manifold. One of the primary advantages is the immediacy and granularity of the data. Unlike traditional market research methods, which often rely on surveys and can be time-consuming and expensive, search engine data is available in real-time and covers a vast range of consumer interactions. This allows companies to monitor market dynamics continuously and respond swiftly to emerging trends and changes in consumer behavior. Moreover, the extensive reach of search engines ensures that the data encompasses diverse demographic and geographic segments, providing comprehensive insights into global market patterns.

However, the analysis of search engine metadata also presents several limitations. One significant challenge is the issue of data privacy and ethical considerations. The use of search data must comply with privacy regulations and ensure that individual users' identities are protected. Another limitation is the potential for noise and irrelevant data within search queries, which can complicate the analysis. Additionally, while search data can indicate interest and intent, it does not always translate directly into consumer action or purchase behavior. Therefore, it is essential to complement search data analysis with other data sources to obtain a more holistic view of the market. Despite these limitations, the strategic use of search engine metadata remains a powerful tool for enhancing market insights and informing business decisions.

Clustering Algorithms in Data Mining

Clustering is a fundamental technique in data mining that involves grouping a set of objects in such a way that objects in the same group (or cluster) are more similar to each other than to those in other groups. This unsupervised learning method is widely used in various fields, including market research, pattern recognition, and image analysis, due to its ability to uncover hidden patterns and structures in large datasets. Several clustering algorithms have been developed over the years, each with its own strengths and applications.

One of the most commonly used clustering techniques is hierarchical clustering, which builds a tree-like structure of nested clusters by either merging smaller clusters into larger ones (agglomerative) or dividing larger clusters into smaller ones (divisive). This method is advantageous for its interpretability and ability to capture the hierarchy of data. Another popular technique is density-based clustering, such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise), which groups points that are closely packed together and identifies outliers as noise. This approach is particularly effective for identifying clusters of arbitrary shapes and handling noise in the data. Other notable clustering algorithms include GMM, which assume that the data is generated from a mixture of several Gaussian distributions, and Spectral Clustering, which uses the eigenvalues of a similarity matrix to perform dimensionality reduction before clustering in fewer dimensions.

Among the various clustering techniques, the K-Means algorithm is one of the most widely used and well-known methods due to its simplicity and efficiency. K-Means clustering aims to partition a dataset into K distinct, non-overlapping clusters, where each data point belongs to the cluster with the nearest mean. The algorithm operates in a straightforward iterative manner: it initializes K centroids randomly, assigns each data point to the nearest centroid, and then updates the centroids to the mean of the assigned points. This process is repeated until the centroids no longer change significantly, indicating convergence.

The popularity of K-Means can be attributed to its computational efficiency, making it suitable for large datasets. Additionally, its simplicity and ease of implementation allow for straightforward application and interpretation of results. However, K-Means also has its limitations. It requires the number of clusters, K, to be specified in advance, which can be challenging without prior knowledge of the data. The algorithm is also sensitive to the initial placement of centroids, which can lead to different results for different initializations. To mitigate this, techniques such as K-Means++ have been developed to improve the initialization process by choosing more spread-out starting points.

Despite these limitations, K-Means remains a powerful tool for data mining, particularly in applications where the goal is to quickly identify and interpret underlying patterns within the data. Its effectiveness in various domains, from customer segmentation in marketing to pattern recognition in image processing, underscores its versatility and enduring relevance in the field of data science.

Applications of K-Means in Marketing

The application of K-Means clustering in marketing has been extensively studied and documented, demonstrating its efficacy in segmenting markets, identifying customer profiles, and enhancing targeted marketing strategies. One notable case study is the work by [18], who used K-Means clustering to segment consumers based on their purchasing behaviors. By analyzing transaction data from a large retail chain, they were able to identify distinct customer segments, such as bargain hunters, brand-loyal shoppers, and high spenders. This segmentation allowed the retailer to tailor marketing efforts to each group, resulting in more personalized promotions and improved customer retention.

Another significant study by [19] applied K-Means clustering to segment customers of a telecommunications company. They used call detail records to categorize users into different groups based on usage patterns, such as heavy users, light users, and occasional callers. The insights gained from this segmentation enabled the company to design targeted marketing campaigns and offer personalized service plans that better matched the needs of each customer group. The success of this approach was reflected in increased customer satisfaction and reduced churn rates.

Methods

To systematically investigate the application of K-Means clustering in enhancing digital marketing strategies through the analysis of flight ticket search patterns, we followed a structured research methodology. This approach ensures a comprehensive analysis, from data collection to the evaluation of clustering results. The methodology is visually summarized in figure 1, which outlines the main steps of our research process.

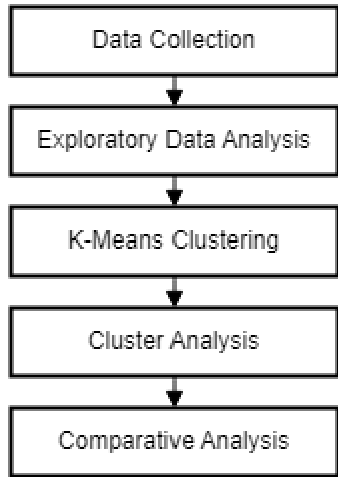

Figure 1 Research Method Flowchart

Data Collection

The dataset utilized in this study consists of search engine results for flight tickets, providing a comprehensive snapshot of user interactions and preferences related to flight searches. Collected on April 1, 2020, the dataset includes a variety of search terms focused on flights to Hong Kong, capturing the top search results and associated metadata. This data offers valuable insights into user behavior and search engine performance, which are crucial for understanding market dynamics in the travel industry. By analyzing this dataset, we aim to uncover patterns and trends that can enhance digital marketing strategies and improve user engagement.

The dataset contains a total of 26 columns, each representing different aspects of the search results. These columns include search terms, ranks, titles, snippets, display links, direct links, query times, total results, cache IDs, formatted URLs, counts, start indices, input encodings, output encodings, safety settings, custom search engine (CX) IDs, geographic locations (GL), search times, formatted search times, and formatted total results. Such a rich collection of metadata allows for a detailed analysis of user search behavior and the effectiveness of various search results.

The key features of the dataset provide a multifaceted view of the search engine results, enabling a thorough analysis of user interactions and preferences. One of the primary features is the search terms, which represent the queries entered by users into the search engine. In this dataset, the search term "flights to Hong Kong" appears repeatedly, indicating a focused interest in this particular route. The rank column specifies the position of each search result, providing insights into the perceived relevance and visibility of the results.

The title and snippet columns offer textual information about each search result. The title typically contains a concise description or headline of the webpage, while the snippet provides a brief summary or excerpt. These textual elements are critical for understanding what information is being presented to users and how it influences their click-through decisions. Additionally, the display link and direct link columns provide the URLs of the search results, indicating the sources and destinations of the web traffic.

Other important features include the query time, which records the exact time when the search was conducted, and the total results, which indicate the overall number of results returned by the search engine for the given query. The cache ID provides a unique identifier for each cached search result, while the formatted URL offers a cleaner, more readable version of the link. The dataset also includes technical details such as input and output encodings, safety settings, and custom search engine IDs, which are essential for understanding the context and configuration of the search.

By leveraging these key features, this study aims to perform a detailed analysis of search engine metadata to segment user search patterns using K-Means clustering. This segmentation will provide actionable insights for enhancing digital marketing strategies, ultimately improving user engagement and satisfaction in the travel industry.

Exploratory Data Analysis (EDA)

The first step in the exploratory data analysis (EDA) process is data cleaning and preprocessing, which is essential for ensuring the quality and reliability of the analysis. Given the diverse nature of the dataset, which includes text and numerical fields, multiple steps were taken to clean and prepare the data. Initially, any missing values in the relevant columns such as search terms, ranks, titles, snippets, display links, and links were addressed by either imputing or removing incomplete records. This step is crucial to prevent any biases or inaccuracies in the subsequent analysis.

Additionally, categorical variables such as search terms and display links were standardized to maintain consistency across the dataset. Textual data in the title and snippet fields were also preprocessed by removing stop words, punctuation, and performing stemming and lemmatization to normalize the text. This preprocessing step enhances the quality of text analysis and clustering by reducing noise and ensuring that similar terms are treated uniformly. These cleaning and preprocessing measures are fundamental to achieving accurate and meaningful insights from the dataset.

One of the key aspects analyzed during the EDA is the distribution of search terms. Understanding the frequency and variety of search terms provides insights into user behavior and preferences. The dataset reveals a concentrated interest in certain search terms, with "flights to Hong Kong" being a prominent example. By visualizing the distribution of these search terms, as shown in figure 2, we can identify the most popular queries and potentially discover emerging trends in user searches.

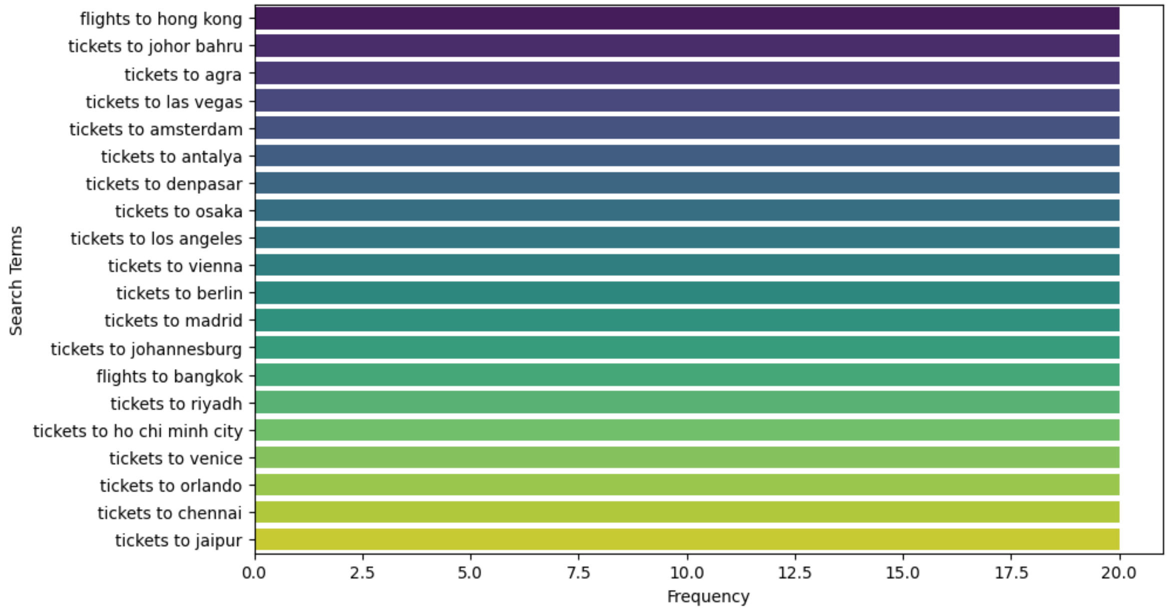

Figure 2 Distribution of Search Terms

The analysis of search term distribution helps in recognizing patterns and focusing marketing efforts on the most relevant terms. For instance, if a significant number of users are searching for flights to a specific destination, targeted marketing campaigns can be developed to cater to this demand. This not only enhances the effectiveness of marketing strategies but also improves user engagement by providing relevant content.

Examining the frequency of keywords within the search terms, titles, and snippets offers further insights into user intent and the relevance of search results. By extracting and analyzing the most common keywords, we can understand what information users are looking for and how it is being presented by various websites. This analysis can be visualized through word clouds or frequency plots, highlighting the most prevalent keywords in the dataset.

The keyword frequency analysis reveals the focus areas of user searches and the effectiveness of the content in attracting clicks. For example, keywords related to "cheap flights" or "discounts" may indicate a price-sensitive user base, prompting marketers to highlight deals and offers in their campaigns. This targeted approach helps in aligning marketing messages with user interests, thereby increasing the likelihood of conversion.

Ranking is a critical factor in search engine results, as it directly influences the visibility and click-through rates (CTR) of the listed websites. The dataset includes a ranking column, which provides valuable information on the position of each search result. Analyzing the distribution of ranks and their correlation with CTR helps in understanding the impact of ranking on user behavior.

Summary statistics for the rank column indicate that the dataset has a mean rank of 5.5, with ranks ranging from 1 to 10. The analysis reveals that higher-ranked results (closer to 1) are more likely to receive clicks compared to lower-ranked ones. This insight underscores the importance of achieving a high rank in search engine results to maximize visibility and user engagement. By focusing on SEO strategies and optimizing content to improve rankings, marketers can enhance the effectiveness of their digital marketing efforts.

K-Means Clustering

K-Means clustering is a popular unsupervised machine learning algorithm used for partitioning a dataset into K distinct, non-overlapping clusters. The algorithm aims to minimize the variance within each cluster, thereby ensuring that the data points within a cluster are as similar as possible, while those in different clusters are as dissimilar as possible. K-Means operates by iteratively assigning each data point to the nearest cluster centroid and then updating the centroids based on the mean of the assigned points. This process repeats until the centroids no longer change significantly, indicating convergence.

The simplicity and efficiency of the K-Means algorithm make it suitable for large datasets and a variety of applications, including market segmentation, pattern recognition, and image analysis. However, the algorithm also has some limitations, such as sensitivity to the initial placement of centroids and the need to specify the number of clusters in advance. Despite these challenges, K-Means remains a widely used and effective clustering technique in data science.

One of the critical steps in applying K-Means clustering is selecting the appropriate parameters, particularly the number of clusters (K). Choosing the right K is crucial as it directly impacts the quality of the clustering results. Several methods can be used to determine the optimal number of clusters, including the Elbow Method, which involves plotting the sum of squared errors (SSE) against different values of K and looking for an "elbow point" where the rate of decrease slows down. This point suggests a balance between the number of clusters and the variance within clusters.

Initialization of centroids is another important parameter that affects the performance of the K-Means algorithm. Poor initialization can lead to suboptimal clustering results or slow convergence. To address this, the K-Means++ initialization technique can be used, which selects initial centroids that are spread out in the data space, improving the chances of finding better cluster assignments. Other parameters that may require tuning include the maximum number of iterations and the tolerance level for convergence.

The first step in implementing K-Means clustering is selecting the features that will be used for clustering. In this study, the features chosen include search terms, titles, and snippets, as they provide meaningful information about user search behavior. These textual features are vectorized using the TfidfVectorizer, which converts the text into numerical values based on the term frequency-inverse document frequency (TF-IDF) representation. This approach ensures that the importance of terms is captured while reducing the influence of commonly occurring words.

Once the features are prepared, the K-Means algorithm is run with the selected parameters. The process begins by initializing the centroids, followed by iteratively assigning data points to the nearest centroid and updating the centroids based on the mean of the assigned points. The algorithm continues this process until the centroids stabilize, indicating convergence. In this study, the Elbow Method was used to determine that the optimal number of clusters is 5. The K-Means algorithm was then executed with K set to 5, providing a clear segmentation of the dataset into distinct clusters.

Evaluating the clustering results is essential to ensure that the algorithm has produced meaningful and useful clusters. Several metrics can be used for this purpose, including the silhouette score, which measures the cohesion and separation of clusters. A high silhouette score indicates that the clusters are well-defined and distinct. In addition, visualizations such as scatter plots of the clustered data can provide intuitive insights into the quality of the clustering. In this study, a PCA was performed to reduce the dimensionality of the data for visualization, and the clusters were plotted to assess their separability and coherence.

By carefully selecting features, running the algorithm with optimal parameters, and thoroughly evaluating the results, K-Means clustering can effectively segment user search patterns, providing valuable insights for enhancing digital marketing strategies. This method enables a deeper understanding of user behavior and preferences, facilitating more targeted and effective marketing efforts.

Results and Discussion

Cluster Analysis

The application of K-Means clustering to the dataset resulted in the identification of five distinct clusters, each representing unique patterns of user behavior and search preferences related to flight tickets. These clusters were derived based on the textual features of search terms, titles, and snippets, vectorized using the TF-IDF method. The optimal number of clusters, determined through the Elbow Method, allowed for a clear segmentation of the dataset, providing valuable insights into the different types of users and their search behaviors.

Each cluster, as shown in figure 3, was analyzed to understand it defining characteristics and the commonalities among the data points within it. The centroids of the clusters, which represent the average position of all the points in a cluster, were examined to identify the prominent features. These centroids provided a concise summary of the main themes and topics associated with each cluster, facilitating the interpretation of the results and the development of targeted marketing strategies.

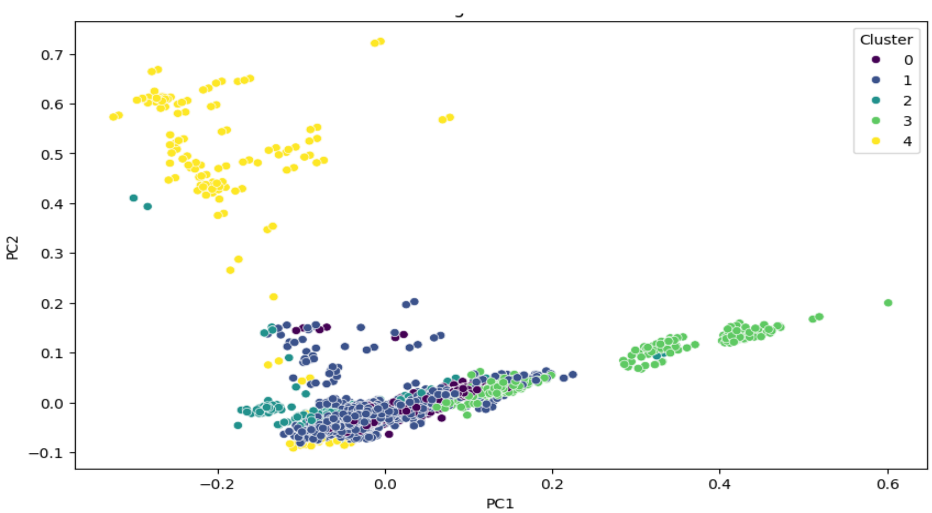

Figure 3 Clustering of Search Terms

The analysis of common search terms within each cluster revealed significant variations in user intent and preferences. For instance, Cluster 1 predominantly featured search terms such as "cheap flights" and "discount tickets," indicating a price-sensitive user segment focused on finding the best deals. In contrast, Cluster 2 included search terms like "premium flights" and "business class," suggesting a user group interested in higher-end travel options. Cluster 3 was characterized by search terms related to specific travel dates and times, highlighting users with fixed travel schedules. Cluster 4's search terms focused on "last-minute flights" and "urgent travel," indicating a segment of users looking for immediate travel solutions. Lastly, Cluster 5 included generic search terms like "flights to Hong Kong," representing users in the initial stages of their travel planning process.

These variations in search terms across clusters underscore the importance of personalized marketing. By understanding the specific interests and needs of each user segment, marketers can tailor their messaging and offers to better resonate with their target audience, thereby enhancing engagement and conversion rates.

In addition to common search terms, the clustering analysis provided insights into the behavioral patterns of users within each segment. Cluster 1, with its focus on budget travel, showed a tendency for users to click on results from discount travel sites and aggregators like Kayak and Skyscanner. These users often compared prices across multiple platforms before making a booking. Cluster 2, on the other hand, displayed a preference for direct bookings through airline websites, particularly for well-known carriers offering premium services. This behavior indicates a trust in brand reputation and a willingness to pay a premium for perceived quality and reliability.

Cluster 3's users demonstrated a pattern of searching for flights well in advance of their travel dates, often seeking flexible ticket options that allow for changes. This group valued convenience and flexibility, likely due to their fixed travel schedules. Cluster 4 exhibited a high sense of urgency, with users frequently engaging with ads and promotions for last-minute deals and rapid booking options. These users prioritized speed and availability over other factors, reflecting their need for immediate travel arrangements.

Finally, Cluster 5 showed a broader exploration pattern, with users engaging with a wide range of search results to gather information about their travel options. This behavior is indicative of users in the early stages of their travel planning, who are still exploring different destinations and flight options.

Impact on Digital Marketing

The insights gained from the K-Means clustering analysis can be directly applied to enhance targeted advertising efforts in the digital marketing domain. By understanding the distinct characteristics and behaviors of each user segment, marketers can design highly specific ad campaigns that resonate with the unique preferences of each cluster. For example, users in Cluster 1, who are primarily interested in cheap flights and discount deals, can be targeted with advertisements highlighting the best flight deals, limited-time offers, and price comparison tools. These ads can be strategically placed on platforms where these users are most active, such as travel deal websites and budget travel forums.

In contrast, users in Cluster 2, who show a preference for premium flights and business class options, can be targeted with advertisements emphasizing luxury travel experiences, exclusive services, and premium airline brands. These ads can be showcased on platforms that cater to business travelers and high-net-worth individuals, such as professional networking sites and premium travel blogs. By tailoring advertising content to match the specific interests of each user segment, marketers can increase the relevance and appeal of their ads, leading to higher click-through rates and better return on investment (ROI).

Personalized offers are another powerful way to leverage the insights from clustering analysis to enhance digital marketing strategies. By segmenting users based on their search behaviors and preferences, travel companies can create customized offers that address the specific needs and desires of each segment. For instance, users in Cluster 3, who prioritize flexibility in their travel plans, can be offered flight tickets with flexible booking options, no change fees, and free cancellations. These personalized offers can be communicated through targeted email campaigns, personalized website content, and tailored mobile app notifications.

Similarly, users in Cluster 4, who are looking for last-minute travel options, can receive personalized offers for urgent travel deals, including flash sales and last-minute discounts. These offers can be delivered through real-time notifications and alerts, ensuring that users receive timely and relevant information that prompts immediate action. By delivering personalized offers that align with the specific needs of each user segment, travel companies can enhance customer satisfaction, foster loyalty, and drive higher conversion rates.

The application of K-Means clustering to segment user search patterns has significant potential to improve engagement and conversion rates in digital marketing. By creating highly targeted and personalized marketing strategies, travel companies can enhance the relevance and effectiveness of their marketing efforts, leading to increased user engagement. When users receive content and offers that closely match their interests and needs, they are more likely to engage with the marketing messages, explore the offered products and services, and ultimately make a purchase.

Moreover, personalized marketing strategies can lead to higher conversion rates by addressing the specific pain points and preferences of each user segment. For example, budget-conscious travelers in Cluster 1 are more likely to convert when presented with compelling deals and discounts, while premium travelers in Cluster 2 are more likely to convert when offered exclusive services and luxury travel experiences. By aligning marketing efforts with the unique characteristics of each user segment, travel companies can optimize their marketing spend, improve the efficiency of their campaigns, and achieve better overall marketing performance.

In summary, the use of K-Means clustering to analyze flight ticket search patterns provides valuable insights that can be used to tailor marketing strategies, enhance targeted advertising, and deliver personalized offers. These strategies not only improve user engagement and satisfaction but also drive higher conversion rates, ultimately contributing to the growth and success of travel companies in a competitive digital marketplace.

Comparative Analysis

Traditional segmentation methods in marketing, such as demographic or geographic segmentation, rely on predefined criteria to group consumers. These methods categorize users based on attributes like age, gender, income, location, and other static characteristics. While useful, these traditional approaches often fail to capture the dynamic and multifaceted nature of consumer behavior. For instance, two individuals of the same age and income bracket may exhibit vastly different search behaviors and preferences when planning their travel. Consequently, traditional segmentation can lead to generalized marketing strategies that may not effectively resonate with all target audiences.

In contrast, K-Means clustering offers a more data-driven and behavior-based approach to segmentation. Instead of relying on predefined categories, K-Means analyzes the actual search behaviors and patterns of users, grouping them into clusters based on similarities in their search terms, titles, snippets, and other features. This method allows for the identification of more nuanced and meaningful segments that reflect real user behavior, enabling more precise and personalized marketing strategies. By focusing on what users do rather than who they are, K-Means clustering provides a deeper understanding of consumer preferences and intentions, which can lead to more effective and targeted marketing efforts.

The advantages of using K-Means clustering for segmentation are numerous. One of the primary benefits is its ability to uncover hidden patterns and relationships within the data that traditional methods might overlook. By grouping users based on their actual behaviors, K-Means can reveal segments that are highly relevant for marketing purposes, such as budget-conscious travelers, last-minute bookers, or luxury seekers. This granularity allows for the development of tailored marketing campaigns that address the specific needs and preferences of each segment, leading to higher engagement and conversion rates.

Another advantage of K-Means clustering is its scalability and efficiency. The algorithm can handle large datasets with relative ease, making it suitable for applications in industries with extensive data, such as travel. Additionally, the iterative nature of the algorithm ensures that the clusters are optimized to minimize variance within each group, enhancing the quality and relevance of the segmentation.

However, K-Means clustering also has its limitations. One significant challenge is the need to specify the number of clusters (K) in advance. Choosing the right number of clusters can be subjective and may require iterative testing and validation methods, such as the Elbow Method, to determine the optimal K. Moreover, K-Means is sensitive to the initial placement of centroids, which can affect the final clustering results. While techniques like K-Means++ can mitigate this issue by improving the initialization process, the sensitivity remains a consideration.

Another limitation is that K-Means assumes clusters are spherical and of similar size, which may not always be the case in real-world data. This assumption can lead to suboptimal clustering when dealing with complex, irregularly shaped data distributions. Despite these limitations, the benefits of K-Means clustering in providing behavior-based segmentation and uncovering meaningful patterns make it a powerful tool for enhancing digital marketing strategies.

Conclusion

The application of K-Means clustering to the search engine metadata for flight tickets provided several key insights into user behavior and preferences. The analysis identified five distinct clusters, each representing unique user segments with specific interests and search patterns. For instance, some clusters were characterized by users searching for budget flights and deals, while others focused on premium travel options or last-minute bookings. These insights reveal the diverse nature of user demands and highlight the importance of tailored marketing strategies to address these varying needs effectively.

Additionally, the clustering analysis demonstrated the value of behavior-based segmentation over traditional demographic methods. By focusing on actual search behaviors, the clusters provided a more accurate representation of user preferences and intentions. This approach enables marketers to develop highly targeted campaigns that resonate with specific user segments, thereby enhancing the relevance and impact of their marketing efforts.

The findings from the K-Means clustering analysis underscore its effectiveness in enhancing digital marketing strategies. The ability to segment users based on their search behaviors allows for the creation of personalized marketing messages and offers, leading to higher engagement and conversion rates. The clusters provided a clear framework for understanding different user groups, enabling marketers to allocate resources more efficiently and design campaigns that cater to the unique needs of each segment.

Moreover, the use of K-Means clustering facilitates real-time decision-making and adaptability in marketing strategies. As user behaviors and market conditions change, the clustering algorithm can be re-applied to update the segments, ensuring that marketing efforts remain relevant and effective. This dynamic and data-driven approach to segmentation is essential for staying competitive in the fast-paced digital marketing landscape.

The insights gained from this study have significant practical applications for the travel industry. Travel companies can leverage the identified clusters to enhance their marketing strategies and improve customer engagement. For example, budget airlines can target the price-sensitive segments with promotions and discounts, while premium airlines can focus on luxury-seeking users with offers that highlight exclusive services and benefits. By aligning marketing efforts with the specific preferences of each user segment, travel companies can increase the effectiveness of their campaigns and drive higher conversion rates.

Furthermore, the clustering results can inform the development of personalized user experiences on travel booking platforms. By recognizing the distinct needs of different user segments, companies can tailor the content, recommendations, and offers displayed to users, enhancing their overall experience and satisfaction. This personalized approach not only improves customer loyalty but also increases the likelihood of repeat bookings and referrals.

Based on the findings, several recommendations can be made for marketers in the travel industry. First, it is essential to adopt a data-driven approach to segmentation, utilizing algorithms like K-Means clustering to gain deeper insights into user behaviors and preferences. Marketers should continuously monitor and analyze search engine data to stay updated on emerging trends and shifts in user demands. This proactive approach enables the development of timely and relevant marketing strategies.

Second, marketers should focus on creating personalized and targeted campaigns that resonate with the specific needs of each user segment. By leveraging the insights from clustering analysis, marketers can design tailored messages and offers that increase user engagement and conversion rates. Additionally, it is crucial to test and optimize these campaigns regularly, using performance metrics to refine and improve their effectiveness.

While this study provides valuable insights into user segmentation using K-Means clustering, there are several areas for future research. One suggestion is to explore the integration of additional data sources, such as social media interactions and user reviews, to enhance the richness and accuracy of the clustering analysis. Combining multiple data streams can provide a more comprehensive understanding of user behaviors and preferences, leading to more precise and effective marketing strategies.

Another area for future research is the application of advanced clustering techniques and algorithms. While K-Means clustering is effective, other methods such as hierarchical clustering, DBSCAN, and Gaussian Mixture Models may offer additional insights and benefits. Comparative studies can evaluate the performance of these algorithms in the context of user segmentation, identifying the most suitable approaches for different marketing scenarios.

To further improve the clustering analysis, future studies could investigate the use of hybrid approaches that combine multiple algorithms. For instance, combining K-Means clustering with hierarchical clustering can enhance the stability and accuracy of the segmentation results. Additionally, machine learning techniques such as deep learning and neural networks can be explored for their potential to uncover more complex patterns and relationships within the data.

Declarations

Author Contributions

Conceptualization: H.T.S. and L.K.O.; Methodology: L.K.O.; Software: H.T.S.; Validation: H.T.S. and L.K.O.; Formal Analysis: H.T.S. and L.K.O.; Investigation: H.T.S.; Resources: L.K.O.; Data Curation: L.K.O.; Writing Original Draft Preparation: H.T.S. and L.K.O.; Writing Review and Editing: L.K.O. and H.T.S.; Visualization: H.T.S.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or

publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. A. Raji, "Digital Marketing in Tourism: A Review of Practices in the Usa and Africa," Int. J. Appl. Res. Soc. Sci., vol. 6, no. 3, pp. 393–408, 2024, doi: 10.51594/ijarss.v6i3.896.

[2] M. Cooper, R. Camprubí, E. Koç, and R. Buckley, "Digital Destination Matching: Practices, Priorities and Predictions," Sustainability, vol. 13, no. 19, p. 10540, 2021, doi: 10.3390/su131910540.

[3] A. Kumar, "Leveraging Digital Innovations in Tourism Marketing: A Study of Destination Promotion Strategies," Int. J. Bus. Manag. Res., vol. 12, no. 1, pp. 08–12, 2024, doi: 10.37391/ijbmr.120102.

[4] S. Afren, "The Role of Digital Marketing Promoting Tourism Business: A Study of Use of the Social Media in Promoting Travel," World J. Adv. Res. Rev., vol. 21, no. 1, pp. 272–287, 2024, doi: 10.30574/wjarr.2024.21.1.2668.

[5] S. Farhangi and H. Alipour, "Social Media as a Catalyst for the Enhancement of Destination Image: Evidence From a Mediterranean Destination With Political Conflict," Sustainability, vol. 13, no. 13, p. 7276, 2021, doi: 10.3390/su13137276.

[6] I. Heliany, "Wonderful Digital Tourism Indonesia Dan Peran Revolusi Industri Dalam Menghadapi Era Ekonomi Digital 5.0," Destin. J. Hosp. Dan Pariwisata, vol. 1, no. 1, pp. 21–35, 2019, doi: 10.31334/jd.v1i1.483.

[7] H. El-Gohary and R. Eid, "DMA Model: Understanding Digital Marketing Adoption and Implementation by Islamic Tourism Organizations," Tour. Anal., vol. 17, no. 4, pp. 523–532, 2012, doi: 10.3727/108354212x13473157390885.

[8] A. Hammad, "Factors Influencing the Digital Tourism Marketing of Generation Z Tourists 'Applied to the Egyptian Tourist Destination,'" Int. J. Tour. Archaeol. Hosp., vol. 4, no. 1, pp. 15–37, 2024, doi: 10.21608/ijtah.2024.251138.1107.

[9] S. Sarjana and J. R. Widokarti, "Strengthening Partnership Strategy for Digital Development in Water Tourism," 2020, doi: 10.2991/aebmr.k.200522.026.

[10] V. Zhezha, "Exploring the Landscape of Digital Marketing in Albania: Insights From Local Companies," Acad. J. Interdiscip. Stud., vol. 12, no. 4, p. 341, 2023, doi: 10.36941/ajis-2023-0120.

[11] F. Aufa, "The Unlocking Natra Bintan's Brand Image: The Power of SEO in Digital Marketing," J. Bisnisman Ris. Bisnis Dan Manaj., vol. 6, no. 1, pp. 108–119, 2024, doi: 10.52005/bisnisman.v6i1.209.

[12] R. Setiawati, “Integrated Digital Payment, Digital Marketing, and Pricing Perception for Decision Purchase of Fruits and Vegetables E-Commerce,” E3s Web Conf., vol. 426, p. 02125, 2023, doi: 10.1051/e3sconf/202342602125.

[13] C. O. Altin, “The Effectiveness of Search Engine Optimization Practices in Promoting Turkey’s Tourism Destinations. A Case Study of Tourism Destinations in Istanbul City,” J. Hosp. Tour. Manag., vol. 6, no. 2, pp. 11–21, 2023, doi: 10.53819/81018102t4137.

[14] R. S. Bhandari and A. K. Bansal, “Impact of Search Engine Optimization as a Marketing Tool,” Jindal J. Bus. Res., vol. 7, no. 1, pp. 23–36, 2018, doi: 10.1177/2278682117754016.

[15] G. Camossi, “Contributions of Search Engine Optimization Techniques to Information Findability,” Encontros Bibli Rev. Eletrônica Bibl. E Ciênc. Informação, vol. 28, 2023, doi: 10.5007/1518-2924.2023.e93371/54069.

[16] H. Choi and H. Varian, “Predicting the Present with Google Trends,” Econ. Rec., vol. 88, no. s1, pp. 2–9, 2012, doi: 10.1111/j.1475-4932.2012.00809.x.

[17] B. Pan, T. MacLaurin, and J. C. Crotts, “Travel Blogs and the Implications for Destination Marketing,” J. Travel Res., vol. 46, no. 1, pp. 35–45, Aug. 2007, doi: 10.1177/0047287507302378.

[18] W. A. Kamakura and R. Y. Du, “How economic contractions and expansions affect expenditure patterns,” J. Consum. Res., vol. 39, no. 2, pp. 229–247, 2012, doi: 10.1086/662611.

[19] C.-Y. Tsai and C.-C. Chiu, “A purchase-based market segmentation methodology,” Expert Syst. Appl., vol. 27, no. 2, pp. 265–276, Aug. 2004, doi: 10.1016/j.eswa.2004.02.005.