Optimizing tourism development using Google Trends and nighttime light (NTL) data

Fitri Kartiasih*, Azhari, Rahadian Eka Bagus Indra Rinangku
[a]*Politeknik Statistika STIS, Jakarta, Indonesia*

Submitted: 22 September 2025 – Revised: 24 November 2025 – Accepted: 03 December 2025

Abstract

Tourism plays a strategic role in driving regional economic growth by creating jobs, preserving culture, and promoting equitable development. Jawa Timur Province recorded the highest number of domestic tourist visits in 2024; however, not all districts and cities have received digital exposure and infrastructure support proportional to their tourism potential. This study analyzes digital visibility, infrastructure development, and tourist arrivals across 38 districts and cities, and classifies them using descriptive analysis, Pearson correlation, and K-Means Clustering. Data sources include Google Trends (digital tourism interest), Nighttime Light imagery (infrastructure development), and domestic tourist statistics from BPS. The results identify four clusters with distinct characteristics: Surabaya City shows high performance across all aspects, while other regions demonstrate strong digital interest but limited infrastructure. The study recommends targeted digital promotion, infrastructure enhancement for high-demand destinations, and strengthened tourism attractiveness in emerging areas, contributing to more efficient, inclusive, and sustainable tourism development.

Keywords: google trends; nighttime light; domestic tourists; clustering
JEL Classification: C22; I18; R11

Recommended Citation
Kartiasih, F., Azhari, & Rinangku, R. E. B. I. (2025). Optimizing tourism development using Google Trends and nighttime light (NTL) data. *Jurnal Ekonomi Indonesia*. 14(3), 252-268. DOI: https://doi.org/10.52813/jei.v14i3.792

Available at: https://jurnal.isei.or.id/index.php/isei

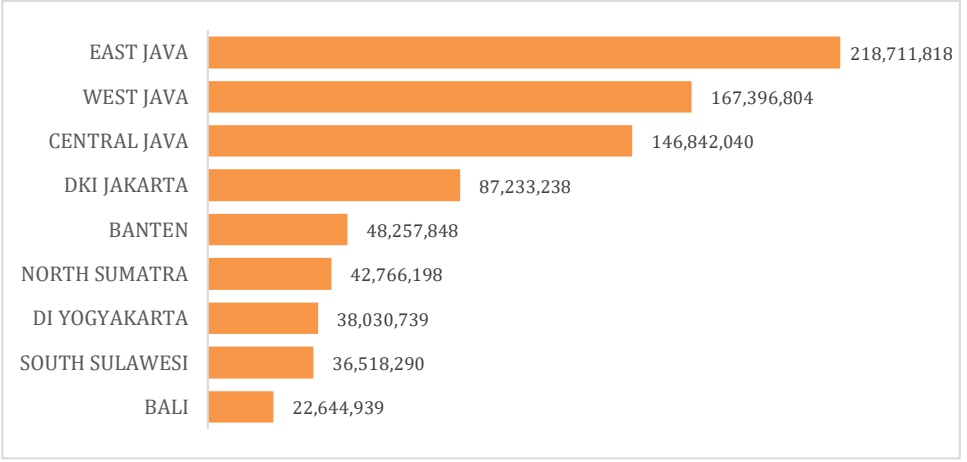
*Corresponding Author: fkartiasih@stis.ac.id

1. Introduction

Tourism is a strategic sector in driving economic growth and regional development (Hafiz et al., 2024). Beyond its contribution to Gross Domestic Product (GDP) and role as a source of foreign exchange, tourism acts as a catalyst for local cultural preservation and job creation. According to the United Nations World Tourism Organization (UNWTO, 2023), this sector contributes approximately 10% to the global economy and is among the fastest to recover in the post-pandemic period. In Indonesia, the tourism sector has demonstrated a strong post-pandemic recovery and has once again become an important contributor to the national economy, highlighting its role in supporting sustainable development.

The digital transformation in travel planning has elevated the importance of online visibility, with tools like Google Trends widely used to map real-time tourism interest (Statista, 2022; Hasyyati et al., 2022). Jawa Timur Province is one of Indonesia's regions with the greatest tourism potential, supported by diverse natural, cultural, historical, and culinary attractions as well as continuously improving infrastructure. In 2024, Jawa Timur recorded the highest number of tourist visits in the country (Figure 1), reaching 218.71 million domestic trips (BPS, 2025). While major cities such as Surabaya, Malang, Batu, and Banyuwangi dominate digital attention, many regions with strong tourism potential, such as Trenggalek, Ngawi, Sampang, and Nganjuk, remain underrepresented online, highlighting an imbalance between actual tourism assets and their digital exposure.

Figure 1.
The 10 Provinces with the Highest Number of Domestic Tourists in Indonesia in 2024



Source: *Badan Pusat Statistik* (2025)

Variations in digital exposure across regions highlight the need for more evidence-based, inclusive tourism promotion: destinations with higher online visibility generally attract more visitors, while those with limited exposure remain less accessible. The Kementerian Pariwisata dan Ekonomi Kreatif (2024) identifies digitalization as a key pillar of destination competitiveness, stressing the importance of data-driven strategies under local budget constraints.

Advances in information technology have expanded access to destination promotion even in remote areas (Torabi et al., 2023), allowing both popular and lesser-known destination promotion even in remote area (Tiago et al., 2021). Nevertheless, tourism performance evaluation in Indonesia faces fundamental methodological challenges. Current analysis relies heavily on conventional statistics gathered through censuses and surveys (Hu et al., 2022). While valid, these methods are often slow, costly, spatially limited (Liu et al., 2022), and less responsive to the rapid changes occurring in the field (Liu et al., 2025)

Beyond digital visibility, infrastructure readiness, such as road access, transportation, public facilities, and digital connectivity, also affects destination competitiveness. However, it is often measured using administrative data that may not reflect real conditions. Nighttime Light Intensity (NTL) offers a more objective indicator of physical development (Elvidge et al., 1997; Shi et al., 2019; Pemayun et al., 2024). However, studies integrating digital data, such as Google Trends, spatial data, such as NTL, and actual visitation figures remain limited, as most Indonesian tourism research still relies on conventional statistics.

To capture regional complexity more effectively, methods that incorporate both spatial and digital variations are needed. Unsupervised techniques like K-Means Clustering can identify regional patterns based on digital interest, infrastructure, and tourist volume. While geospatial big data offers a fast and adaptive representation of tourism activity (Wei et al., 2022), traditional datasets often lack the spatial detail needed for real-time visualization (Xu et al., 2021), underscoring the need for more integrative analytical approaches. Consequently, this study aims to bridge this gap by employing machine learning to objectively segment regions of integrating these three distinct data sources.

Therefore, this study aims to: (1) describe the characteristics of tourism interest and infrastructure in Jawa Timur districts and cities; (2) cluster these regions using the K-Means method based on Google Trends, NTL, and actual visitation data; and (3) provide data-driven policy recommendations for equitable tourism development. This study contributes by integrating these three distinct data sources and adopting a machine learning approach for objective segmentation, thereby supporting evidence-based policymaking for more targeted and efficient resource allocation in Jawa Timur.

2. Methodology

The data used in this study were compiled from multiple reputable sources. Official statistics on the number of domestic tourist trips by district and city in Jawa Timur Province were obtained from Indonesia's Central Statistics Agency (BPS). In addition, data on public search interest were sourced from Google Trends, with the regional filter set to Indonesia to ensure national-level relevance.

The composite Nighttime Light (NTL) index was derived from the Visible Infrared Imaging Radiometer Suite (VIIRS) instrument on the Suomi National Polar-orbiting Partnership (Suomi-NPP) satellite and accessed via Google Earth Engine (GEE).

The analysis period covers data from 2024. This study adopts a quantitative approach by utilizing official statistics and big data to analyze and identify regions based on the number of domestic tourist trips, digital exposure, and infrastructure availability as supporting factors.

The data collection and processing were carried out using Google Earth Engine (GEE), R-Studio, Google Colab, and QGIS. The research variables used are as follows:

Table 1.
Research Variables

Variable	Unit	Source
Number of Domestic Tourists	Trips	BPS – Statistics
Google Trends Index	Scale [0,100]	Google
Nighttime Light	$nW/cm^2/sr$	Suomi-NPP

Source: Processed by Author

Domestic tourists who travel within the territory of a country, in this case Indonesia, for a duration of less than six months and not for the purpose of earning income at the destination. Such travel is also not considered a routine trip (such as commuting for school or work), but rather to visit commercial tourist attractions and/or stay in commercial accommodations, and/or involves a travel distance of at least 100 (one hundred) kilometres round trip (BPS and Kementerian Parisiwata, 2017).

His definition of domestic tourists has also been adopted by BPS based on the international definition developed by the WTO, which states that domestics tourists must meet the following criteria: 1) Indonesia residents; 2) Trip duration < 6 months; 3) Travel within Indonesia; 4) Voluntary/involuntary travel; 5) Not for work or school purposes; 6) Visiting Tourist Attractions; 7) Staying in commercial accommodations; 8) Travel distance ≥ 100 km round trip.

Google Trends is a service that provides data and statistical charts on the popularity of search topics on the internet, based on users' geographic regions and languages (Kamal et al., 2024). The service presents information in the form of an index ranging from 0 to 100 (Jane, et. al., 2024). This score is generated by comparing the search volume of a particular term (search query) against the total search volume within a given region and time period. According to Google Support, the data available in Google Trends are divided into two categories: real-time data, which consist of random samples of searches from the past seven days prior to when the trend data are generated, and non–real-time data, which consist of samples of searches from 2004 up to approximately 36 hours before the trend data are generated.

This research began with a data collection and preprocessing stage. The Google Trend Index (GTI) was constructed using a list of tourist destinations in Jawa Timur obtained from the Jawa Timur Provincial Department of Culture and Tourism website. Data retrieval was conducted for each district/city by identifying the top five search keywords in each region. The average value of these five keywords was used to represent the GTI for each district/city.

Nighttime light, or luminosity, is a composite index derived from satellite imagery that is widely used to analyze the socio-economic conditions of a region. NTL data reflect the level of human activity and infrastructure development occurring at night, thereby serving as an indirect representation of economic and developmental dynamics (Kamal et. al., 2024).

Areas showing higher levels of nighttime illumination generally have more intense economic activities and more advanced development and infrastructure. Therefore,

NTL is considered relevant for measuring various aspects of development, including infrastructure.

NTL data serve as a proxy for assessing the level and availability of tourism infrastructure spatially by comparing them with tourist visit data to identify areas with imbalances or inadequacies between infrastructure and tourism interest (Gao et al., 2024). The Nighttime Light (NTL) data were obtained from the Google Earth Engine platform, with preprocessing involving the selection of satellite imagery with the lowest cloud coverage within one year. The imagery was then processed using the zonal statistics method, aggregating mean values at the district/city level

To achieve the research objectives, two analytical methods were applied: descriptive and inferential analysis. The descriptive analysis involved presenting summary statistics for each variable under study to provide an initial overview of the data distribution and characteristics. The inferential analysis was carried out using Pearson correlation to identify the strength and direction of the relationships between domestic tourist trips, digital exposure, and infrastructure availability and clustering analysis was then performed to group districts/cities based on similarities across the three variables.

The Pearson analysis technique is used to measure the strength and direction of the linear relationship between two variables measured on an interval or ratio scale. The Pearson correlation coefficient is calculated using the following formula (Schober, 2018):

$$r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} \quad (1)$$

Cluster analysis using the non-hierarchical method (K-Means) is applied to classify individuals or objects with similar characteristics, allowing for a more detailed analysis of the differences between groups (Kartiasih et. al., 2025). The partitioning K-Means method was employed for clustering, with the Elbow method used to determine the optimal number of clusters. The Elbow method was employed to determine the optimal number of clusters (Maori and Evanita, 2023). The clustering results were subsequently utilized as the basis for data-driven and targeted tourism policy recommendations. We applied the non-hierarchical K-Means clustering method to classify the regions (Ahmar et. al., 2018), utilizing Euclidean Distance to measure the similarity between objects.

3. Results and Discussion

To examine the potential of tourism interest and infrastructure across districts and cities in Jawa Timur Province in 2024, summary statistics of the three main variables used in the analysis are presented, namely the Google Trends Index (GTI), Nighttime Light (NTL), and the number of domestic tourists, as shown in Table 2.

Table 2.
Summary Statistics of Variables Used

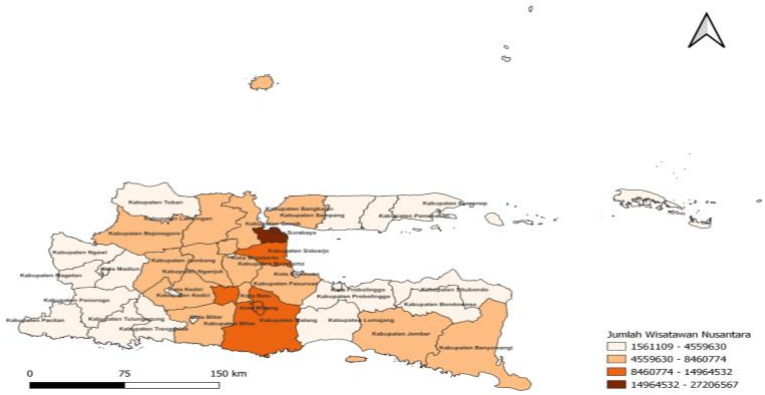
Statistic	Number of domestic tourists	GTI	NTL
Minimum	1,561,109	1.89	0.74
Median	4,353,216	13.87	1.99
Mean	5,755,574	17.21	4.19
Maximum	27,206,567	58.63	21.15
Kurtosis	2.85	1.28	4.18
Skewness	10.29	1.16	2.10

Source: Processed by Author

The number of domestic tourists serves as the primary variable representing the level of domestic travel to each district and city in Jawa Timur Province. A striking disparity can be observed in the number of visits, ranging from 1,561,109 in Mojokerto City to more than 27 million in Surabaya City. The average number of visits is 5,755,574, with a median of 4,353,216, indicating substantial variation across districts and cities. Regions with exceptionally high numbers of visitors tend to act as major tourism magnets, while many other areas remain lagging in attracting domestic tourists.

The Google Trends Index (GTI) represents the level of public search interest in tourism-related keywords on Google, serving as an early indicator of public interest in a region as a tourist destination. The GTI values range from 1.89 percent to 58.63 percent, with an average of 17.21 percent and a median of 13.87 percent. This indicates that most districts and cities in Jawa Timur Province remain below the average level of public interest, although several areas, such as Malang Regency, Surabaya City, Pasuruan Regency, Pacitan Regency, Batu City, and Malang City exhibit particularly high search levels, positioning them as potential hubs of tourism growth in the province. The ability of GTI to serve as an early indicator of tourism interest has been widely acknowledged, as digital platforms have become the primary channel for travel planning and destination discovery (Hasyyati et al., 2022; Statista, 2022).

Figure 2.
Distribution of Domestic Tourist by Regency/City in Est Java Province in 2024

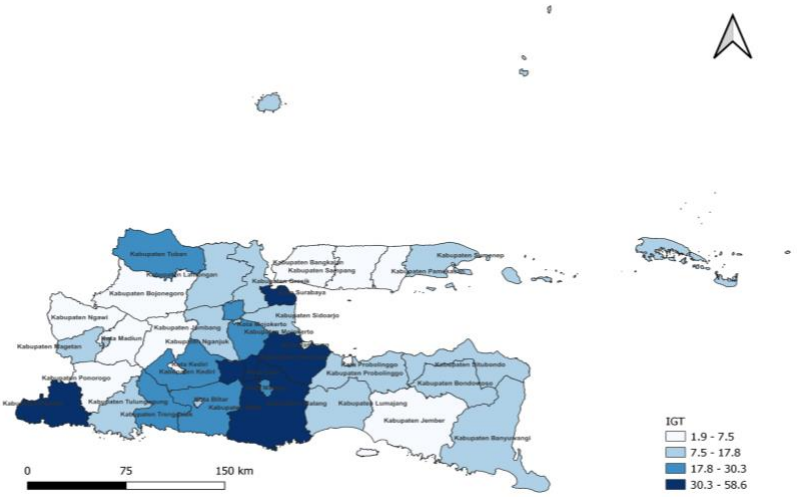


Source: Processed by Author

Nighttime Light (NTL) reflects the level of nighttime brightness, which is closely associated with the degree of infrastructure development and economic activity. In the districts and cities of Jawa Timur Province, NTL values range from a minimum of 0.74 to a maximum of 21.15, with an average of 4.19 and a median of 1.99.

The higher mean compared to the median indicates the presence of several areas with exceptionally high levels of nighttime illumination, found in almost all urban areas of the province such as Surabaya City, Madiun City, Mojokerto City, Malang City, and others, as well as Sidoarjo Regency, which functions as a buffer area for Surabaya. Thus, NTL can serve as a supporting indicator for identifying regions with facilities and infrastructure that are more conducive to tourism activities

Figure 3.
Overview of the Google Trends Index (GTI) Distribution by Regency/City in Jawa Timur Province in 2024



Source: Processed by Author

Overall, the three variables provide an initial overview that the potential and development of tourism in Jawa Timur Province remain uneven across districts and cities. Some areas appear well-recognized and capable of attracting significant public attention and large numbers of tourists, while others are still in the development stage.

The results of the Pearson correlation analysis show a strong positive relationship between the number of domestic tourists and the Google Trends Index (GTI) at 0.62, as well as a moderate positive relationship between the number of domestic tourists and Nighttime Light (NTL) at 0.42 in Jawa Timur Province in 2024. This strong correlation between digital search interest and actual tourist arrivals aligns with seminal research demonstrating the superior predictive power of online search data for tourism demand forecasting (Bangwayo-Skeete & Skeete, 2015). Meanwhile, the moderate positive relationship with NTL confirms that infrastructure development, as a proxy for economic activity, is a significant determinant of tourism attractiveness. This supports its use as an indirect measure of tourism readiness, an approach validated by foundational studies that have used nighttime light imagery to accurately

measure economic growth from outer space (Henderson et al., 2012).

Figure 4.
Overview of the Nighttime Light (NTL) Distribution by Regency/City in Jawa Timur Province in 2024

Source: Processed by Author

These findings indicate that an increase in public interest in tourist destinations, as reflected in Google search intensity, tends to be followed by an increase in the number of domestic tourist visits. In addition, higher levels of economic activity and infrastructure quality, as represented by NTL, also contribute to the attractiveness of a region for tourists.

Table 3.
Correlation of GTI and NTL with the Number of Domestic Tourists in 2024

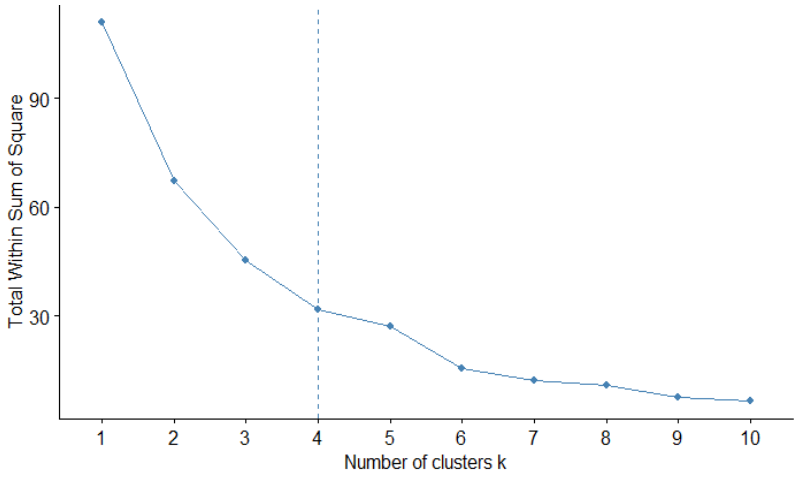
Variable	Correlation
GTI	0.62
NTL	0.42

Source: Processed by Author

Therefore, Google Trends can serve as an early indicator for monitoring domestic tourism behavior. At the same time, NTL reflects the importance of infrastructure development in supporting the tourism sector in Jawa Timur Province. In this study, we employed a clustering method to group districts/cities using the non-hierarchical K-Means approach. Before clustering, the optimal number of clusters was determined using the Elbow method. The optimal value of *k* is identified from the Elbow plot at the point where the slope of the curve starts to level off, indicating diminishing improvements in within-cluster variance reduction.

Figure 5.

Elbow Graph



Source: Processed by Author

As shown in Figure 5, the "elbow point" in the plot suggests an optimal k-value of four, where the slope of the curve begins to level off. Subsequently, the clustering process using the K-Means method with four clusters was performed, and the results are presented in Figure 6.

Table 4.
Mean of Variable for Each Cluster

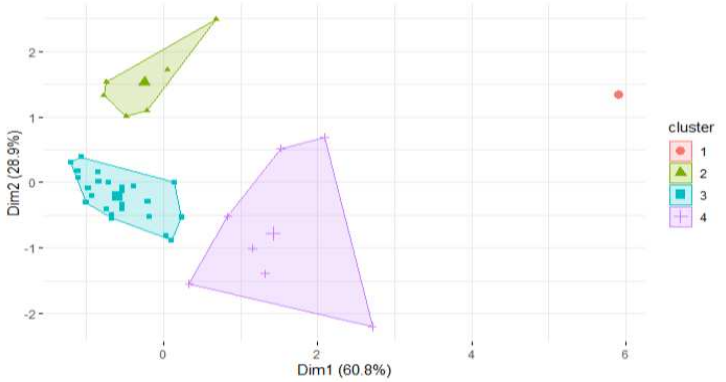
Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4
GTI	49.23	8.33	12.31	37.04
NTL	21.15	10.74	1.76	4.50
Number of domestic tourists	27,206,567	2,258,217	4,544,158	9,842,308

Source: Processed by Author

Subsequently, the districts/cities in Jawa Timur Province were grouped into four clusters, with the criteria for each cluster determined through profiling. The classification of regions was established based on the average values of each variable in Table 4 relative to the assigned cluster labels, as detailed below.

Cluster 1 comprises regions with very high interest in tourism destination searches, a large number of tourist visits, and highly supportive infrastructure. Surabaya City is the only region classified in Cluster 1, characterized by a strong public interest in searching for its tourism destinations, an exceptionally high number of tourist visits, and well-developed supporting infrastructure. As the capital city of Jawa Timur Province and the centre of regional economic activity, Surabaya offers a wide range of attractions, including historical, cultural, and culinary tourism, as well as modern leisure options such as urban parks and shopping centres.

Figure 6.
K-Means Clustering Visualization



Source: Processed by Author

Figure 7.
Grouping of Regencies/Cities in Jawa Timur Province in 2024

Source: Processed by Author

The city also excels in providing tourism-supporting facilities, including transportation access, accommodation networks, public services, and digital connectivity that facilitates easy access to tourism information. Cluster 2 consists of less popular regions characterized by low search interest and a low number of domestic tourists, yet supported by adequate infrastructure. Kediri City, Blitar City, Probolinggo City, Pasuruan City, Mojokerto City, and Madiun City fall into the group

of regions with relatively low levels of tourism-related search interest and domestic tourist visits. Yet, they already possess fairly well-developed tourism-supporting infrastructure.

This indicates that although destinations in these areas are not yet widely searched for or visited in large numbers, they are physically and logistically prepared to accommodate tourists. Such readiness may include the availability of good road networks, adequate public facilities, accessible transportation options, and accommodation and other basic services.

Cluster 3 consists of moderately popular regions that exhibit moderate levels of search interest and domestic tourist visits, but have underdeveloped infrastructure. A total of 24 regencies/cities fall into Cluster 3, characterized by moderate levels of tourism-related search interest on Google and a moderate number of domestic tourist visits, yet lacking optimally supportive infrastructure. This indicates that the regions within this cluster face challenges, including limited numbers of well-established tourist destinations and low physical readiness to support tourism activities. Infrastructure limitations may include inadequate transportation access, insufficient accommodation facilities, a lack of sanitation and security services, and underdeveloped tourism-supporting amenities.

Cluster 4 comprises popular regions that show high search interest and a large number of domestic tourist visits, supported by moderately developed infrastructure. This cluster consists of seven areas, namely Pacitan Regency, Malang Regency, Pasuruan Regency, Sidoarjo Regency, Mojokerto Regency, Malang City, and Batu City. These areas are characterized by high levels of tourism-related search interest and a large number of domestic tourist visits, indicating that their destinations are well-known and attractive to the public. This reflects the existence of strong market potential in the tourism sector, both in terms of destination appeal and public awareness. Nevertheless, these regions still require improvements in the availability and quality of tourism-supporting infrastructure to fully optimize their potential.

The regional grouping from the K-Means analysis identifies several categories of districts/cities in Jawa Timur Province were identified, each with similar characteristics in terms of tourism potential and development. This clustering provides an essential foundation for formulating more targeted and data-driven policy recommendations. Each group of regions reflects different conditions and needs; therefore, the follow-up actions must be tailored to the profile of each cluster. The following recommendations may be considered by the Jawa Timur Provincial Government, particularly through the Department of Culture and Tourism and other relevant agencies.

Cluster 1 represents regions with very high tourism search interest and visitor numbers, supported by strong infrastructure. The high level of interest in Surabaya's tourism destinations is reflected not only in the large number of domestic tourist visits but also in the high frequency of online searches. Keyword searches for Surabaya's attractions reveal that the most frequently searched destinations include the Surabaya Museum, G-Walk, Surabaya North Quay, Surabaya Zoo, and Wonorejo Mangrove. These findings indicate that tourists' interests in Surabaya are diverse, encompassing educational, recreational, and nature-based attractions.

The Surabaya Museum reflects tourists' interest in history and educational values. G-Walk and Surabaya North Quay highlight the appeal of modern tourism, offering entertainment, culinary experiences, and vibrant nightlife. Meanwhile, Surabaya Zoo is a popular choice for family-oriented visits and educational experiences on

biodiversity. The Wonorejo Mangrove is attracting growing interest in nature and ecotourism. Overall, these preferences illustrate that Surabaya is capable of offering a wide range of tourism experiences that cater to the diverse needs and interests of visitors.

This positions Surabaya as a model of an urban area whose tourism management has reached a mature stage. However, the challenge for Cluster 1 lies not merely in increasing visitor numbers, but in managing tourist flows and ensuring sustainability. As indicated by urban tourism studies, cities such as Surabaya face risks related to congestion and environmental pressure (Iswari et al., 2025). The strategic focus should therefore shift toward innovative tourism solutions leveraging high digital connectivity to manage crowds, promote off-peak travel, and prioritize high-value experiences rather than sheer visitor volume.

Future development should focus on enhancing the quality of the visitor experience, diversifying community-based tourism products, and strengthening the creative economy sector as a key driver of sustainable tourism. Amid challenges of urbanization and environmental pressures, approaches such as smart tourism and green infrastructure present significant opportunities for Surabaya to maintain its appeal as one of Jawa Timur's flagship destinations.

Thus, Surabaya's inclusion in Cluster 1 underscores its position as a metropolitan city that excels not only in economic performance and governance but also in building a strong and sustainable tourism image. The combination of strong digital presence, high visitor numbers, and robust infrastructure makes it an ideal example for other regions seeking to enhance the overall competitiveness of their tourism sector.

Cluster 2 consists of less popular regions characterized by low search interest and a low number of domestic tourist visits, despite being supported by adequate infrastructure**.** This cluster highlights a gap between the supply of infrastructure and the level of tourism demand.

The infrastructure that has already been developed has not yet been fully utilized or effectively linked to strong tourism potential and digital promotion strategies. The low level of online searches for tourism destinations also shapes tourists' perceptions of the attractiveness of these areas. Yet, many of these cities possess rich historical, cultural, and local assets that could be leveraged as unique selling points for tourism development.

At the same time, this situation offers strategic opportunities for tourism growth through enhanced destination appeal and more effective promotion. The primary strategy should be the implementation of aggressive digital promotion. Referring to Awad and Alharthi (2025), these regions need to invest in Electronic Word of Mouth (e-WOM) and social media marketing to convert existing infrastructure readiness into actual tourism demand. Urban infrastructure can be utilized to host festivals, sports events, and other activities that generate organic digital content, while tourism narratives should shift from administrative functions to distinctive urban attractions.

Keyword searches for Cluster 2 indicate public interest in entertainment, recreational, and educational tourism, as reflected in destinations such as Sunrise Mall Mojokerto and Suncity Waterpark. Supported by relatively well-developed infrastructure, future development should prioritize strengthening destination identity through integrated digital promotion, thematic tour packages, and inter-regional collaboration.

In sum, this cluster represents regions with strong growth potential if their tourism assets are creatively and innovatively enhanced. The existing infrastructure serves as

an important foundation, but it must be complemented by efforts to strengthen destination content and digital connectivity. If optimized, these areas could successfully transition from being perceived as "less popular" destinations into attractive alternative tourism hubs supported by their already available infrastructure.

Cluster 3 comprises moderately popular regions that exhibit medium levels of search interest and domestic tourist visits but are constrained by underdeveloped infrastructure**.** Cluster 3 illustrates that tourism in these regions is still in the early stages of development or that the tourism sector has not yet been positioned as a priority within local development agendas. The existing tourism potential is likely not fully recognized, promoted, or packaged in an appealing manner. In addition, the absence of systematic promotion and the limited use of digital narratives have further marginalized these regions from the mainstream flow of tourist movement. Low visitation numbers may also stem from tourists' perceptions of restricted accessibility, convenience, and available facilities.

Nevertheless, being classified in Cluster 3 should not be seen solely as a weakness, but rather as a strategic area for medium- to long-term tourism development. Developing these regions requires a grassroots-oriented approach. Large-scale infrastructure projects may not be immediately economically viable; instead, the "Tourism Village" model, supported by Pentahelix collaboration among government, academia, business, community, and media stakeholders, is essential (Noviana et al., 2025).

Research from Batu City further indicates that community participation and local regulations are more critical than massive physical infrastructure during the early stages of rural tourism development. An initial step would be to map local potential based on natural, cultural, and social resources that reflect the unique identity of these regions. Development of basic infrastructure should also be pursued in parallel with community engagement, ensuring that tourism growth addresses not only attractions but also the social and economic readiness of the local population. Cross-sector partnerships among local governments, private stakeholders, academia, and communities will be key to establishing a sustainable foundation for tourism.

In this regard, Cluster 3 comprises regions with high investment needs for tourism development and capacity-building. While the challenges are considerable, the potential for growth remains substantial if development strategies are well-targeted. Moving forward, these areas could be prioritized in provincial tourism equalization programs, not only to create new destinations but also to strengthen inclusivity and fairness in the distribution of economic benefits from the tourism sector.

Cluster 4 includes popular regions that demonstrate high search interest and a large number of domestic tourist visits, supported by moderately developed infrastructure. Cluster 4 highlights the need for infrastructure enhancement to optimize the tourism potential of these regions. The destinations driving high search interest in Cluster 4 are predominantly nature-based, including Mount Bromo, Taman Beach in Pacitan, Coban Rondo Waterfall, Batu Paragliding, and Hawai Waterpark Malang.

This indicates that the majority of tourists are attracted to nature-based destinations in these areas. Although these destinations already possess high levels of appeal, providing optimal comfort and convenience for visitors requires further infrastructure development to keep pace with increasing tourist arrivals. Without such improvements and adjustments, there is a risk of declining visitor satisfaction,

which could ultimately affect tourist loyalty and the long-term reputation of these destinations.

The presence of regions within this cluster underscores the urgency of accelerating infrastructure development. In contrast to Cluster 2, which primarily requires marketing interventions, Cluster 4 demands substantial investment in infrastructure and destination management. Sholeha and Sumarmi (2025) emphasize the need for a *Coastal Ecosystem Services* approach in areas such as Pacitan (Taman/Perawan Beach) to manage the environmental impacts of high visitor numbers on fragile ecosystems.

Accordingly, policy efforts should prioritize conservation, waste management, and accessibility improvements without compromising the natural character of the destinations. With existing tourism demand already established, investments in improving accessibility, providing standard tourism facilities, and integrating destinations are likely to have a significant positive impact on the visitor experience. Furthermore, strengthening local institutional capacity and encouraging community involvement are equally important to ensure that development remains inclusive and sustainable.

In other words, Cluster 4 represents regions that are already in a promising stage of tourism growth but still require interventions in supporting aspects to maximize their potential. Infrastructure improvements in these areas will not only enhance tourist satisfaction but also contribute to local economic growth, job creation, and balanced regional development.

This study's primary strength lies in its innovative methodological approach, which integrates big data, such as Google Trends, and spatial data from Nighttime Light (NTL). This fusion, analyzed through K-Means clustering, provides a promotional and infrastructural gap. Consequently, the research makes a significant contribution to evidence-based policymaking, offering a foundation for developing more targeted, efficient, and balanced tourism strategies at the district or city level. This data-centric model serves as a powerful tool for transforming regional tourism planning from intuition-based to evidence-led.

However, the research is subject to several weaknesses and limitations. Its heavy reliance on digital and spatial data sources means the analysis is constrained by the scope and potential biases of this data, risking the omission of crucial human or cultural factors that are not easily quantified. The dynamic nature of tourism also limits the study; its predictions, based on historical data, may lose accuracy due to sudden shifts in tourist behaviour influenced by economic conditions of new policies, necessitating continuous validation and adaptation.

4. Conclusion

The findings of this study conclude that the number of domestic tourists, the Google Trends index, and Nighttime Light (NTL) data can provide valuable insights into the potential and development of the tourism sector in Jawa Timur Province. However, the analysis indicates that such development remains uneven across districts and municipalities. Using the K-Means Clustering method, regions were grouped according to indicators of search interest in tourist destinations, the number of domestic tourist visits, and the level of infrastructure support.

The results show that districts and municipalities are divided into four clusters. Cluster 1 represents areas with very high levels of destination search interest and tourist visits, supported by adequate infrastructure, specifically, the City of Surabaya. Cluster 2 consists of six cities with low popularity and minimal tourist visits, despite having relatively well-developed infrastructure.

Cluster 3 comprises 23 districts/municipalities with moderate levels of search interest and tourist visits, but limited infrastructure support. Meanwhile, Cluster 4 includes five districts and two cities with high search interest and tourist visits, alongside moderately sufficient infrastructure support. Policy recommendations were formulated according to the distinct characteristics of each cluster optimize the development of tourism potential in these areas.

References

[1] Ahmar, A.S., Napitupulu, D., Rahim, R., Hidayat, R., Sonatha, Y. and Azmi, M. (2018) 'Using K-means clustering to cluster provinces in Indonesia', *Journal of Physics: Conference Series*, 1028(1), p. 012006. doi:10.1088/1742-6596/1028/1/012006.

[2] Awad, A., & Alharthi, B. (2025). *The role of digital marketing tools in promoting tourism: An applied study on online marketing strategies*. Innovative Marketing, 21(2), 14–26. https://doi.org/10.21511/im.21(2).2025.02

[3] BPS (Badan Pusat Statistik) (2025) *Statistik Pariwisata Indonesia 2024*. Jakarta: BPS RI.

[4] BPS (Badan Pusat Statistik) and Kementrian Pariwisata (2017) *Kajian Data Pasar Wisatawan Nusantara*. Jakarta: BPS RI.

[5] Elvidge, C.D., Baugh, K.E., Kihn, E.A., Kroehl, H.W. and Davis, E.R. (1997) 'Mapping city lights with nighttime data from the DMSP Operational Linescan System', *Photogrammetric Engineering and Remote Sensing*, 63(6), pp. 727–734.

[6] Gao, Z. et al. (2024) 'Exploring tourist spatiotemporal behavior differences and tourism infrastructure supply–demand pattern fusing social media and nighttime light remote sensing data', *International Journal of Digital Earth*, 17(1). doi:10.1080/17538947.2024.2310723.

[7] Hafiz, M.S., Ismail, G., Wilda, M.S. and Kartiasih, F. (2024). The Effect of Exchange Rate Volatility on Foreign Tourist Visits in Indonesia: GARCH Analysis. *Jurnal Ekonomi dan Bisnis*, 13(3), pp.421-433.

[8] Hasyyati, U., Herlambang, T. and Ardiansyah, R. (2022) 'Analisis tren digital pariwisata menggunakan Google Trends dan media sosial', *Jurnal Teknologi dan Sistem Informasi*, 3(2), pp. 89–98.

[9] Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994–1028.

[10] Hu, Y., Zhang, Z., Li, Z. and Li, X. (2022) 'Using survey data to analyze spatial-temporal patterns in tourism: A review', *Tourism Management Perspectives*, 41, p. 100958. doi:10.1016/j.tmp.2021.100958.

[11] Iswari, H. R., Baits Nur, A. W. M. R., & Rusdianti, I. S. (2025). *Developing Malang City tourism: A qualitative analysis of opportunities, challenges, and sustainability strategies*. Pangripta: Jurnal Ilmiah Kajian Perencanaan Pembangunan, **8**(2), 160–178.

[12] Jane, G.J., Hasabi, R., Purnatadya, S.D. and Kartiasih, F. (2024). Utilizing Google Trends Data to Examine the Impact of Open Unemployment Rates on Indonesia's Gross Domestic Product. *Sistemasi: Jurnal Sistem Informasi*, 13(6), pp.2299-2320.

[13] Kamal, F.Y., Sari, M.I., Utami, M.F.G.U. and Kartiasih, F. (2024). Penggunaan Remote Sensing dan Google Trends untuk Estimasi Produk Domestik Bruto Indonesia. *Equilibrium: Jurnal Penelitian Pendidikan dan Ekonomi*, 21(02), pp.37-59.

[14] Kartiasih, F., Azhari, and Rinangku, R.E.B.I. (2025). Carbon Trading Potential in the Forest Sector: Analysis at the Province Level in Indonesia. *EKO-REGIONAL: Jurnal Pembangunan*

Ekonomi Wilayah, 20(1), pp.14-29.
[15] Kementerian Pariwisata Republik Indonesia (2024) *Siaran Pers: Kemenparekraf dukung akselerasi transformasi digital pelaku parekraf lewat BEDIL*, Kementerian Pariwisata RI, 12 November. Available at: https://www.kemenpar.go.id/berita/siaran-pers-kemenparekraf-dukung-akselerasi-transformasi-digital-pelaku-parekraf-lewat-bedil (Accessed: 11 September 2025).
[16] Liu, X., Zhang, Y. and Gao, S. (2022) 'The limitations of traditional tourism statistics and the rise of spatial big data', *Annals of Tourism Research*, 92, p. 103316. doi:10.1016/j.annals.2021.103316.
[17] Liu, X., Zhang, Y., Wu, C. and Gao, J. (2025) 'Real-time tourism monitoring with big data: Challenges and future directions', *Journal of Destination Marketing & Management*, 26, p. 100779. doi:10.1016/j.jdmm.2024.100779.
[18] Maori, N.A. and Evanita, S. (2023) 'Metode Elbow dalam optimasi jumlah cluster pada K-Means Clustering', *Jurnal SIMETRIS*, 14(2), pp. 277–287.
[19] Noviana, L., Situmorang, M. T. N., Cahyaputri, B., Farida, F., & Sukwika, T. (2025). *Analysis of Pentahelix Tourism Village for Ecotourism Development in Batu City, East Java*. International Journal of Sustainable Development and Planning, 20(3), 1273–1282. https://doi.org/10.18280/ijsdp.200332
[20] Pemayun, A.A.G.R.B.D., Azizi, M.Z., Daulay, N.A., Apriliani, N.H. and Kartiasih, F. (2024). Estimation of Java GRDP in Regency/City Level: Satellite Imagery and Machine Learning Approaches. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 10(2), pp.379-386.
[21] Schober, P., Boer, C. and Schwarte, L.A. (2018) 'Correlation coefficients: Appropriate use and interpretation', *Anesthesia & Analgesia*, 126(5), pp. 1763–1768. doi:10.1213/ANE.0000000000002864.
[22] Shi, T. (2019) 'Spatial data mining and big data analysis of tourist travel behavior', *IIETA Conference Proceedings*.
[23] Sholeha, A. W., & Sumarmi, S. (2025). *Managing sustainable tourism of Perawan Beach, East Java, Indonesia based on the Coastal Ecosystem Service (CES) approach*. Geojournal of Tourism and Geosites, 58(1), 78–88. https://doi.org/10.30892/gtg.58107-1392
[24] Statista (2022) *Leading travel-related search topics on Google worldwide*. Available at: https://www.statista.com/statistics/1124484/google-search-volume-tourism-worldwide/ (Accessed: 10 September 2025).
[25] Tiago, M.T.P.M.B., Amaral, F., Tiago, F. and Veríssimo, J.M.C. (2021) 'Digital marketing and social media: Why bother with destination branding?', *Journal of Business Research*, 129, pp. 902–911. doi:10.1016/j.jbusres.2019.09.031.
[26] Torabi, F., Ghaderi, J. and Dolnicar, S. (2023) 'Technology and remote destinations: The role of digital infrastructure in tourism development', *Tourism Management*, 96, p. 104699. doi:10.1016/j.tourman.2022.104699.
[27] UNWTO (United Nations World Tourism Organization) (2023) *Tourism and COVID-19: Impacts and Future Perspectives*. Madrid: UNWTO.
[28] Wei, L., Zhong, Y. and Fan, W. (2022) 'Mapping tourism activities using geo-spatial big data: A new framework', *Computers, Environment and Urban Systems*, 92, p. 101743. doi:10.1016/j.compenvurbsys.2021.101743.
[29] Xu, H., Zhang, Y. and Liu, Y. (2021) 'Integrating spatial data into tourism statistics: Opportunities and challenges', *Tourism Geographies*, 23(4), pp. 705–723. doi:10.1080/14616688.2020.1745019.

Appendix 1. Clusters of Districts/Cities with Tourism Potential in Jawa Timur Province

District/City	Cluster
Kota Surabaya	1
Kota Kediri	2
Kota Blitar	2
Kota Probolinggo	2
Kota Pasuruan	2
Kota Mojokerto	2
Kota Madiun	2
Kabupaten Ponorogo	3
Kabupaten Trenggalek	3
Kabupaten Tulungagung	3
Kabupaten Blitar	3
Kabupaten Kediri	3
Kabupaten Lumajang	3
Kabupaten Jember	3
Kabupaten Banyuwangi	3
Kabupaten Bondowoso	3
Kabupaten Situbondo	3
Kabupaten Probolinggo	3
Kabupaten Jombang	3
Kabupaten Nganjuk	3
Kabupaten Madiun	3
Kabupaten Magetan	3
Kabupaten Ngawi	3
Kabupaten Bojonegoro	3
Kabupaten Tuban	3
Kabupaten Lamongan	3
Kabupaten Gresik	3
Kabupaten Bangkalan	3
Kabupaten Sampang	3
Kabupaten Pamekasan	3
Kabupaten Sumenep	3
Kabupaten Pacitan	4
Kabupaten Malang	4
Kabupaten Pasuruan	4
Kabupaten Sidoarjo	4
Kabupaten Mojokerto	4
Kota Malang	4
Kota Batu	4