Sustainable Development Research; Vol. 7, No. 2; 2025
ISSN 2690-9898 E-ISSN 2690-9901
https://doi.org/10.30560/sdr.v7n2p35

Investigation on Tourism Trends Using K-means Clustering and Regression Analysis

Anna Sheila Ilumin Crisostomo[1*], Badar Al Dhuhli[1] & Reggie C. Gustilo[2]

[1] Department of Tourism and Management, Oman Tourism College, Muscat, Sultanate of Oman

[2] Department of Electronics and Computer Engineering, De La Salle University, Manila, Philippines

Correspondence: Anna Sheila Ilumin Crisostomo, Department of Tourism and Management, Oman Tourism College, Muscat, Sultanate of Oman. E-mail: anna.crisostomo@otc.edu.om[1*], badar.al-dhuhli@otc.edu.om[1], reggie.gustilo@dlsu.edu.ph[2]

Received: March 26, 2025 Accepted: May 8, 2025 Online Published: May 8, 2025

Abstract

Purpose: The purpose of this study was to analyze tourism trends by determining the clusters of tourists based on common factors. Three (3) characteristics were explored using k-means clustering namely tourists' demographics, travel patterns and travel preferences. These clusters were based on individual's age, gender, country of origin, frequency of travel, travel destinations and seasons. Regression analysis was also performed to determine the factors that influence the length of stay of tourists in their travel destinations.

Methodology: This research conducted a survey from 150 respondents of different age groups, gender, and nationalities. Frequency of travel in a year, length of stay per travel, seasons, destinations, purpose of travel and preferred booking method were the parameters inquired in the survey. The collected dataset was utilized to characterize the clusters of tourists with common considerations. Additionally, regression analysis was used to forecast predictors influencing tourists' length of stay.

Findings: Three (3) parameters were considered in performing k-means clustering such as tourists' demographic profiles, travel patterns and preferences. Regression analysis likewise was employed to predict visitors' length of stay using age, gender, purpose of travel, travel season, and preferred destination as independent variables. In participants' demographics, number of clusters generated was k=5. Gender and nationalities were found to be randomly clustered while other parameters were categorized according to various age groups and frequency of travel. Consequently, for tourists' travel patterns, age, gender, country of origin, frequency of stay, purpose of travel, length of stay and travel seasons were used as parameters. The elbow method knee-point revealed (k=6) as the optimal number of clusters. Moreover, travel preferences parameter was also considered for clustering where predictors like gender, age, country of origin, frequency of travel, purpose of travel, travel season and length of stay were utilized. The optimal number of clusters for this category generated K=5. Regression analysis revealed gender, age and purpose of travel as significant factors influencing tourists' average length of stay. The combination of these variables generated the lowest value of MSE=0.64.

Research limitations/implications: A limited dataset of 150 respondents mainly from Asia and Middle East were utilized in performing preliminary initiatives in analyzing tourism trends. The predictors used in the analysis were restricted to gender, age, country of origin, travel frequency, length of stay, travel season and travel destinations. Supplementary parameters ca be considered in a big data setting for similar studies in the future. K-means clustering was selected among other algorithms with attributes commonality while regression analysis was employed to determine the factors influencing tourists' length of stay in their destinations.

Social Implications: Results of this study will greatly support individual tourists in determining trends in various travel destinations. Similarly, business owners gain benefit forecasting travellers' requirements such as accommodation, food, services, etc. Research findings likewise provide informed decisions for stakeholders

Originality / Value: The dataset used were participants from different countries and nationalities which include Philippines, Saudi Arabia, United Arab Emirates, Oman, USA, Portugal, Germany, Malaysia, Thailand, Qatar, Finland, Denmark, Spain Taiwan, South Korea, Singapore, Australia, Austria, England, UK, India and China. The presented codes were programmed in python where analyses and interpretations were based on formulated objectives. K-means clustering, and regression analysis were both employed to present varied clusters according to tourists' demographic profiles, travel patterns and preferences. Different factors were identified and used to predict tourists' length of stay in their preferred destinations.

Keywords: tourism trends, k-means clustering, regression analysis, forecasting

1. Introduction

Tourism is one of the major economic drivers of a country. It serves as a significant source of revenue which fosters economic growth, promotes cultural exchange, and creates diverse employment opportunities. Capitalizing on tourism results to various beneficial outcomes such as attracting more investment prospects, global reputation enhancement and local community development. Once tourism trends are determined, understood, and analysed, policymakers, business owners, researchers and even individuals can benefit by recognizing possible trends and travellers' demands and preferences. These are also enablers to make informed decisions and maximize tourism sector's potential as well as reveal valuable insights and strategies for stakeholders to adopt and further innovate current offerings. In this digital era, tourism data is increasingly accessible through surveys and online platforms which can be utilized to its maximum. Patterns can be analysed and predicted utilizing demographic trends, destinations' attractiveness and reputation, tourists' behaviour, and overall economic impact. Analysis of this data assists destinations enhancing their attractions, attracting diverse tourist groups, and effectively compete in the global tourism market. Additionally, stakeholders will be able to draft apt strategies meeting travellers' expectations aligned with sustainable tourism practices.

1.1 Problem Statement

The ongoing increase in tourism demand triggers the need to further understand the trends and requirements of individual travellers. Determining the tourism patterns will facilitate local authorities, business owners and tourists in categorizing and quantifying the behaviour and clusters of tourists. This research aims to explore tourists' characteristics according to their demographic information, travel patterns and preferences using k-means clustering. Similarly, it would likewise be interesting and beneficial to know the parameters that influence the length of stay of tourists in their preferred destinations.

1.2 Research Questions

- What are the tourist groups and characteristics according to their demographic profiles, travel patterns, and preferences using k-means clustering?
- What are the parameters that influence travellers' length of stay in their preferred destinations?

1.3 Research Objectives

- To examine and characterize tourist groups according to their demographic profiles, travel patterns and preferences using k-means clustering.
- To determine and analyse the parameters that influence tourists' length of stay in their preferred destinations.

2. Literature Review

The integration of K-means clustering, and regression analysis offers robust insights into tourism trends. Clustering is widely used for segmenting tourists based on behavioral patterns and preferences. Yildirim, Kaya, & FurkanInce (2022) utilized k-means clustering algorithm to obtain key parameters of profiles clustered into groups according to their characteristics. Findings revealed that frequency of tourist vacations, the time between bookings and vacations, and age are the most significant parameters for a tourists' profile. Another study by Hasanah, Sudibyo, and Galih (2021) used K-means clustering algorithm to perform data classification on tourists' country of origin and employed silhouette score to determine the applicable number of clusters. Various types of tourist attractions in Bali were clustered according to their popularity using K-Means and X-Means algorithms (Monica, Natalia, and Sudirman, 2018).

Furthermore, a research initiative by Wang et al. (2018) investigated tourist behavior patterns in urban destinations by clustering geotagged social media data. The researchers used K-means to group tourists based on their activity preferences, such as cultural attractions, nightlife, and shopping hubs. Similarly, Zhou and Chen (2021) used geotagged Instagram information to analyze tourist movement patterns between attractions.

To examine the relationship between destination image dimensions, tourists' previous visits and whether tourists are domestic or international as well as their intention to revisit heritage destination, ordinal logistic regression was employed (Kaur and Kaur, 2019). Meanwhile, Brida and Scuderi (2013) explored the determinants of tourist expenditure using regression analysis. Results revealed that income and length of stay were the most significant predictors for expenditure. Similarly, Peng et al., (2015) employed meta regression analysis to investigate the correlation between estimated international tourism demand elasticities and the data attributes and study features that may impact these empirical estimates.

Ruiz-Reina (2021) applied regression and clustering in the hotel accommodation market. Another study addressing tourist flow prediction challenges utilized multi-feature combined prediction model, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise-Incremental Entropy-K-means clustering-Time Convolutional Network-Coati Optimization Algorithm (CEEMDAN-IEK-TT-COA) (Feng et al., 2024). A combination of clustering and forecasting algorithms were employed in a research performed by Li et al. (2020). This experiment utilized K-means algorithm noting seasonal clustering and a tourist flow prediction method particle swarm optimization-least squares support vector machine (PSO-LSSVM) algorithm to forecast the tourist flow in tourist destinations. Another study applied a combination of tourism forecasting model using an artificial neural network (ANN) and a clustering algorithm, which includes two parameters of the data series: sequence patterns and near characteristics (Jun et al., 2018).

Research conducted by Bartl et al. (2025) employed fuzzy clustering to determine distinct tourist types including respondents' diverse behavior while additive logistic regression analysis was likewise utilized to analyze temporal changes in travel behavior patterns. Findings indicate that the chance of belonging to distinct tourist types changes over tourists' age, time (period) because of external aspects and across generations (cohort). Additionally, Aguilar and Diaz (2019) employed survival analysis approach to analyze the degree the tourists' socio demographic profiles and travel characteristics determine the length of stay. It was found out that country of origin, destination, time, and budgetary restrictions were the most significant factors. Li, et al. (2016) also utilized cluster-based logistic regression model for the holiday travel choice prediction which revealed the essential impacts of influencing factors (gender, driver's license status, the number of car household, household income, number of elders and children traveling together, travel time and stay time) on travel mode selection.

3. Research Methodology

The main objective of this research was to explore and investigate tourism trends and patterns to assist and support stakeholders' efficient planning, implement policies, and improve tourist destinations in the accommodation and other sectors of tourism. Figure 1 presents the three (3) phases adopted for the successful completion of the research.

Figure 1. Three Phases of Study for Analysing Tourism Trends

3.1 Data Gathering

Collection of data was carried out from the designed online survey according to the objectives' requirements. A total of 150 respondents were accumulated from various age groups, gender, and tourists' preferences. The collected data covered critical aspects, including:

- Demographic details of tourists (e.g., age, gender, nationality).
- Travel behaviors (e.g., destination choice, length of stay, travel budgets).
- External factors namely, seasonal trends, destination attractiveness, and economic indicators such as income levels and exchange rates.

3.2 K-means Clustering for Tourism Segments

Subsequently, the study applies K-means clustering, a machine learning technique, to group tourists into distinct segments based on shared characteristics. This step supports in identifying meaningful patterns, such as:

- Different types of tourists (e.g., adventure seekers, luxury travellers, or budget-conscious visitors).
- Popular destinations and activities that appeal to each group.

- Seasonal or regional trends in tourist behaviors.

Elbow Method or Silhouette Analysis was employed to determine the ideal number of clusters. Each cluster represents a specific group of tourists with similar preferences to understand their unique requirements.

3.3 Regression Analysis to Understand Key Drivers

Regression analysis was then performed to explore the factors influencing key tourism trends such as tourists' length of stay, travel spending as well as external elements namely, destinations' reputation, affordability, and availability of amenities.

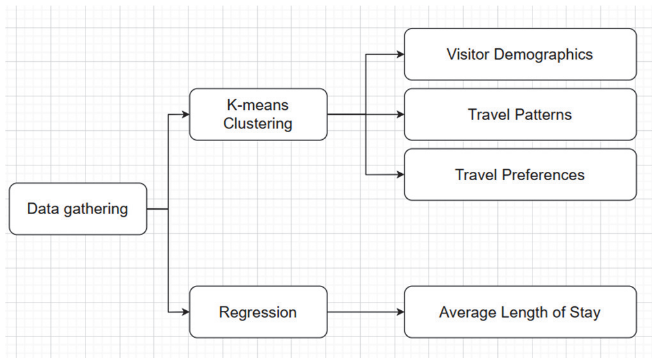

Figure 2. Analysis of Tourism Trends using K-means Clustering and Linear Regression

4. Results

The following sections are the results of implementing k-means clustering and regression analysis to analyse tourism trends.

4.1 K-means Clustering

4.1.1 Tourists' Demographic Profiles

Multiple demographic parameters such as age, gender, country of origin, and annual frequency of travel were utilized to group respondents through k-means clustering. The initial analysis was carried out identifying the optimal number of k (groups) to describe and analyse the segmentation of respondents. The elbow-point method was used to determine the actual number of groups and to detect if increasing its value significantly improves the performance of clustering. Figure 3 below shows the result of the elbow-point method where k=5 presented the optimal number of clusters to appropriately analyse visitors' demographics.

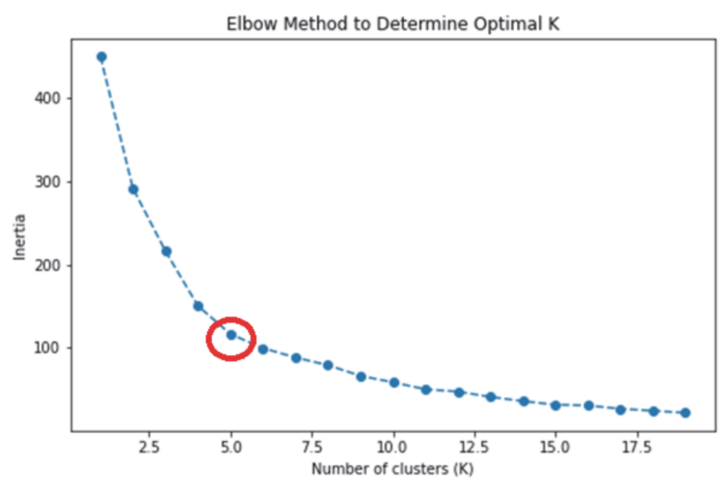

Figure 3. Elbow-point method for Tourists' Demographics

Since there were more than two (2) predictors used in k-means clustering for tourists' demographic profiles, the principal component analysis (PCA) was utilized to exhibit respondents' clustering.

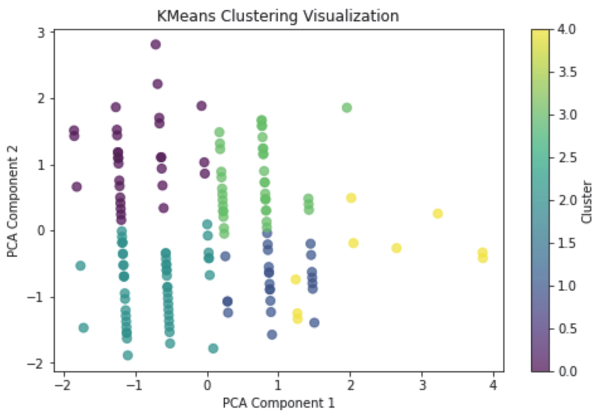

Figure 4. K-means clustering Result Using PCA

Table 1. Tourists' Demographic Results using K-means Clustering

Group	Details
1	Average age: 26.64 years old Gender: Random Nationality: Random Average frequency of travel: 1.82 times per year
2	Average age: 55.58 years old Gender: Random Nationality: Random Average frequency of travel: 1.61 times per year
3	Average age: 43.95 years old Gender: Random Nationality: Random Average frequency of travel: 1.95 times per year
4	Average age: 33.83 years old Gender: Random Nationality: Random Average frequency of travel: 2 times per year
5	Average age: 39.83 years old Gender: Random Nationality: Random Average frequency of travel: 2.19 times per year

Table 1 above presents the analysis of each group according to various factors. Findings were generated after performing K-means clustering algorithm. Each group is characterized by their average age, gender distribution, nationality, and average frequency of travel per year. While gender and nationality are categorized as random in this dataset, the notable differences in age and travel frequency provide insights into their distinct travel patterns and potential preferences.

4.1.1.2 Group 1: The Younger, Occasional Travelers

This group consists of young individuals with an average age of 26.64 years and 1.82 trips per year suggesting occasional travel preference as compared with other clusters. This behaviour might be influenced by financial constraints, early career stages, or academic commitments. This particular age group may have prioritized budget-friendly travel options and adventure-focused destinations.

4.1.1.3 Group 2: Older, Less Frequent Travelers

Average age of 55.58 years represents older individuals, possibly pre-retirees or retirees. They have the lowest average travel frequency among all groups at 1.61 trips per year. This could reflect a more selective approach to travel, prioritizing comfort and value over frequency. Health considerations, lifestyle preferences, or financial stability may also shape their travel behaviour. Destinations offering relaxation, cultural enrichment, or high-quality amenities might be more appealing to this group.

4.1.1.4 Group 3: Middle-Aged and Moderately Active Travelers

This group embodies middle-aged individuals, with an average age of 43.95 years, and exhibits a moderate travel frequency of 1.95 trips per year. These individuals may be in the peak of their careers, balancing family, work, and leisure. Their travel behaviour suggests a preference for planned vacations that fit into their structured schedules and budget. Destinations offering a mix of family-friendly options and personal enrichment opportunities could resonate well with this group.

4.1.1.5 Group 4: Young Adults with a Higher Travel Frequency

Category 4 has an average age of 33.83 years and a travel frequency of 2 trips per year, classifying them among the more frequent travellers. This group likely consists of young professionals who have established financial independence, allowing for more frequent leisure trips. They might seek a balance between relaxation and adventure while favouring destinations that offer varied social experiences. Their relatively high travel frequency suggests they may also value weekend getaways or shorter trips.

4.1.1.6 Group 5: Experienced and Frequent Travellers

The group with an average of 39.83 years is classified as the most frequent travellers, averaging 2.19 trips annually. These individuals may have greater financial capacity or lifestyle that prioritizes travel. This cluster might have established career, lesser responsibilities, or have strong personal interest in exploration. This group may favour diverse and high-value destinations, ranging from cultural landmarks to luxury retreats, as their travel preferences might reflect both experience and enthusiasm for discovery.

4.1.2 Comparative Insights

Age Trends: Travel frequency does not directly correlate with age. For instance, while Group 2 is categorized as the oldest, they travel the least. Meanwhile Group 5, despite being older than Groups 1 and 4, travels the most. This suggests that factors like financial stability, life stage, and priorities might influence travel behaviour more than age.

Travel Frequency: Clusters 4 and 5 exhibit the highest travel frequencies, indicating more priority is given to travel as a significant part of their lifestyle. Conversely, Groups 1 and 2 travel less frequently, most likely due to limited resources or other life priorities.

Potential Preferences: The younger groups (1 and 4) might lean towards budget-friendly and adventurous destinations, while older groups (2 and 5) may prefer more relaxed, high-quality, or cultural experiences. Group 3, being in the middle age range, might seek family-oriented or mixed-purpose travel options.

4.1.3 Travel Patterns

Tourists' travel patterns were also explored using age, gender, country of origin, frequency of stay, purpose of travel, length of stay and travel season as predictors. The elbow method was likewise utilized to determine the optimal number of groups for this category.

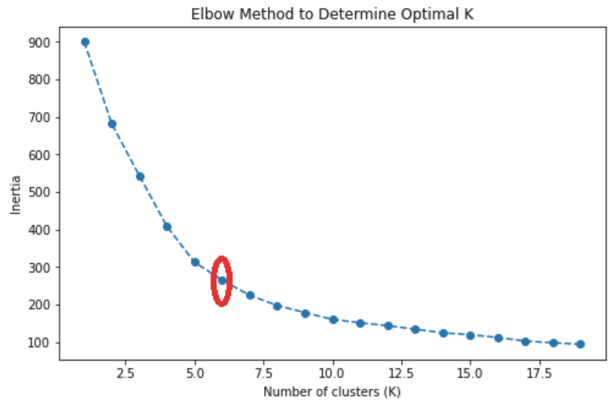

Figure 5. Elbow-point Method Result for Tourist Patterns

Figure 5 illustrates the knee-point or the optimal number of clusters (k=6) to accurately analyse travel patterns of distinct individuals. Using PCA, figure 6 reflects the groupings of the tourists' travel patterns.

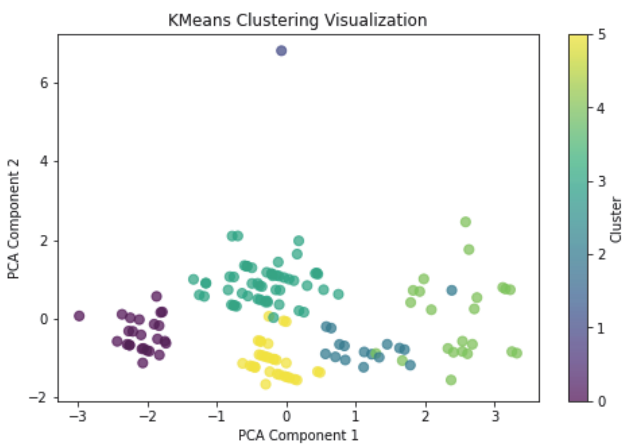

Figure 6. Result of k-means clustering for Travel Patterns

Table 2 below demonstrates the results and analysis of k-means clustering individual-wise components for travel patterns.

Table 2. Travel Patterns Details Group-Wise

Group	Details
0	Average age: 39.26 years old Frequency of travel: 2.16 times Purpose of travel: Either Business or Religious Travel season: Spring or Winter
1	Average age: 31.5 years old Frequency of travel: 1.65 times Purpose of travel: Either Leisure and recreation or Education Travel season: Either Winter or Autumn
2	Average age: 32.69 years old Frequency of travel: 1.88 times Purpose of travel: Mostly Leisure and Recreation

	Travel season: Summer
3	Average age: 41.38 years old Frequency of travel: 2.08 times Purpose of travel: Education Travel season: Summer
4	Average age: 38 years old Frequency of travel: 20 times Purpose of travel: Either Business or Religious Travel season: Summer
5	Average age: 56 years old Frequency of travel: 1.44 times Purpose of travel: Mostly Leisure and Recreation Travel season: Either Summer or Spring

The dataset presents insights on the six distinct groups of travellers, segmented based on their average age, frequency of travel, purpose of travel, and preferred travel season. By examining these variables, patterns are identified reflecting motivations and behaviours of each group, providing a foundation for targeted tourism planning and marketing strategies.

4.1.3.1 Group 0: Middle-Aged Professionals and Devotees

- Average age: 39.26 years
- Frequency of travel: 2.16 times per year
- Purpose of travel: Either business or religious
- Travel season: Spring or winter

Group 0 represents middle-aged individuals who moderately travelled, with a frequency of just over two trips annually. Their travel purpose is often professional (business-related) or spiritual (religious commitments), reflecting a more purpose-driven approach to travel. Their seasonal preference for spring and winter suggests they may prioritize cooler climates or align their travel with specific events, such as business conferences or religious pilgrimages. This group likely seeks efficient travel options and well-organized itineraries.

4.1.3.2 Group 1: Younger Travelers Seeking Leisure or Education

- Average age: 31.5 years
- Frequency of travel: 1.65 times per year
- Purpose of travel: Either leisure and recreation or education
- Travel season: Either winter or autumn

This group is characterized by younger individuals with a lesser travel frequency, averaging 1.65 trips per year. Their primary motivations for travel are leisure, recreation, or education, indicating a mix of relaxation and personal development. The seasonal preference for winter or autumn could be associated to school calendars or opportunities to enjoy off-peak travel periods. This group is likely to favour destinations offering a balance of affordability, cultural experiences, and opportunities for learning or personal enrichment.

4.1.3.3 Group 2: Summer Enthusiasts Focused on Leisure

- Average age: 32.69 years
- Frequency of travel: 1.88 times per year
- Purpose of travel: Mostly leisure and recreation
- Travel season: Summer

Group 2 consists of travellers who prioritize leisure and recreation, with a moderate travel frequency of 1.88 times per year. Their preference for summer season indicates they enjoy warmer climate and outdoor activities, more to adventure tourism. This group is likely to seek destinations known for vibrant summer attractions, including festivals, scenic landscapes, and recreational opportunities. They may also be drawn to family-friendly travel options during the summer holiday season.

4.1.3.4 Group 3: Educators and Summer Travelers

- Average age: 41.38 years
- Frequency of travel: 2.08 times per year
- Purpose of travel: Education
- Travel season: Summer

This group is slightly older than Group 2, with a similar travel frequency of just over two trips per year. Their primary purpose of travel is education, suggesting that they may include professionals attending training, conferences, or workshops. Their preference for summer aligns with academic schedules and suggests an emphasis on career development or knowledge-sharing opportunities. This group may require destinations with educational facilities, convenient accommodation options, and tailored travel services.

4.1.3.5 Group 4: High-Frequency Travellers for Business and Religion

- Average age: 38 years
- Frequency of travel: 20 times per year
- Purpose of travel: Either business or religious
- Travel season: Summer

Group 4 stands out with an exceptionally high travel frequency, averaging 20 trips annually. Their primary purposes are business and religious, indicating they are likely professionals or individuals engaged in regular religious obligations. Their preference for summer suggests consistent commitments during this season, such as annual conferences or religious events. This group likely values reliability, efficiency, and flexibility in travel arrangements, as well as access to amenities that cater to their specific demands.

4.1.3.6 Group 5: Older Leisure Travelers

- Average age: 56 years
- Frequency of travel: 1.44 times per year
- Purpose of travel: Mostly leisure and recreation
- Travel season: Either summer or spring

As the oldest group in the dataset, Group 5 reflects travellers who primarily focus on leisure and recreation. Their lower travel frequency, averaging 1.44 trips per year, may stem from financial, health, or lifestyle factors. Their preference for summer or spring suggests they enjoy mild weather and scenic destinations, making them likely candidates for cultural tours, nature-based travel, or relaxed vacation experiences. This group may prioritize comfort, accessibility, and destinations with slower-paced activities.

4.1.4 Comparative Insights

Age and Frequency of Travel: While age does not directly correlate with travel frequency, older groups (e.g., Group 5) tend to travel less frequently, except for Group 4, which shows unusually high travel rates likely due to specific obligations. Younger groups generally show moderate travel frequency, driven by leisure or educational purposes.

Purpose of Travel: Business and religious travels dominate Groups 0 and 4, while leisure and education define Groups 1, 2, 3, and 5. The motivations reveal clear distinctions, such as education-focused summer trips for Group 3 versus high-frequency religious/business travel for Group 4.

Seasonal Preferences: Summer is the most popular season across groups, reflecting its appeal for vacations and professional gatherings. Groups 0 and 1 also highlight winter and autumn as travel seasons, possibly for off-peak travel or specific purposes like education or relaxation.

4.1.5 Travel Preferences

Gender, age, country of origin, travel frequency, purpose of travel, travel season and length of stay were identified as predictors for clustering tourists' travel preferences. Elbow-point method generated an optimal (k=5) as reflected in figure 7.

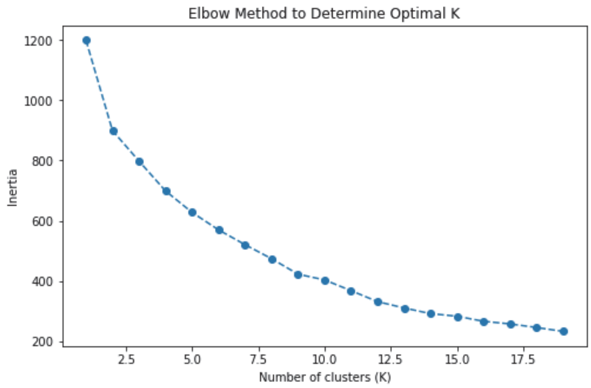

Figure 7. Result of Elbow-point method for Travel Preferences

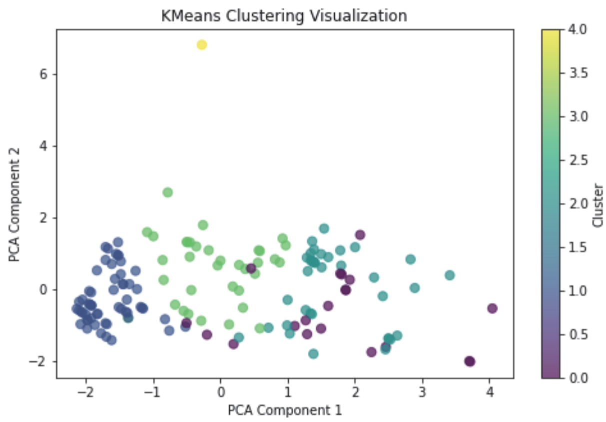

Figure 8. K-means clustering for Travel Preferences

The interpretation for the groupings or clusters for Travel Preferences is shown in table 3.

Table 3. Clustering Details for Travel Preferences

Group	Details
1	Average age: 40.74 years old Preferred travel season: Spring Preferred travel destinations: Cities and Mountains Booking method: Direct or Online
2	Average age: 68 years old Preferred travel season: Spring Preferred travel destination: Cultural Sites Booking method: Direct
3	Average age: 27.7 years old Preferred travel season: Spring Preferred travel destinations: Cities and Mountains Booking method: Direct or Travel agency
4	Average age: 55.69 years old Preferred travel seasons: Spring and Summer Preferred travel destinations: Cities and Mountains Booking method: Direct or Online
5	Average age: 39.56 years old

Preferred travel seasons: Spring and Summer
Preferred travel destinations: Beach and Nature
Booking method: Direct

The data highlights five distinct traveller groups, characterized by their average age, preferred travel season, favourite destinations, and booking methods. These insights offer a comprehensive understanding of the preferences and habits of each group, which can be valuable for tourism service providers and marketers aiming to cater to specific traveller needs.

4.1.5.1 Group 1: Middle-Aged Urban and Mountain Travelers

- Average age: 40.74 years
- Preferred travel season: Spring
- Preferred travel destinations: Cities and Mountains
- Booking method: Direct or Online

Group 1 consists of middle-aged individuals who favour spring for travel, possibly drawn to the mild weather and blooming landscapes. Their preference for urban and mountain destinations suggests they enjoy a mix of cultural exploration and outdoor activities. This group values flexibility in booking, as both direct and online options suit their needs. These travellers likely appreciate well-planned itineraries that combine city tours with scenic mountain excursions.

4.1.5.2 Group 2: Older Cultural Travelers

- Average age: 68 years
- Preferred travel season: Spring
- Preferred travel destination: Cultural Sites
- Booking method: Direct

Group 2 represents the oldest demographic in the dataset, with a strong preference for cultural destinations such as museums, historical landmarks, and heritage sites. Their travel aligns with the spring season, which often features pleasant weather and cultural festivals. This group tends to book their trips directly, possibly indicating a preference for personal interaction and clear communication when arranging their travels. Destinations targeting this group should focus on cultural enrichment and accessibility to meet their needs.

4.1.5.3 Group 3: Young Explorers of Cities and Mountains

- Average age: 27.7 years
- Preferred travel season: Spring
- Preferred travel destinations: Cities and Mountains
- Booking method: Direct or Travel Agency

Group 3 consists of younger travellers who share love for cities and mountains, similar to Group 1. However, they often rely on travel agencies in addition to direct booking. Their preference for spring suggests an interest in vibrant, refreshing experiences. This group may seek affordable and adventurous travel options, such as guided mountain hikes or city tours with social elements. Travel agencies catering to this group could offer customized packages that emphasize exploration and affordability.

4.1.5.4 Group 4: Mature Travellers with Diverse Preferences

- Average age: 55.69 years
- Preferred travel seasons: Spring and Summer
- Preferred travel destinations: Cities and Mountains
- Booking method: Direct or Online

Group 4 consists of mature travellers who enjoy a mix of city exploration and mountain adventures, with flexibility in their preferred travel seasons. Their interest spans spring and summer, suggesting value for both blooming landscapes of spring and the vibrant energy of summer. They show a balanced approach to booking, leveraging both direct and online platforms. This group likely seeks travel experiences that blend comfort with moderate adventure, making them ideal for premium city tours and scenic retreats.

4.1.5.5 Group 5: Nature and Beach Enthusiasts

- Average age: 39.56 years
- Preferred travel seasons: Spring and Summer
- Preferred travel destinations: Beach and Nature
- Booking method: Direct

Group 5 travellers are nature lovers who favour beaches and serene landscapes, reflecting their desire for relaxation and outdoor beauty. With a preference for spring and summer, this group aligns their travels with seasons ideal for enjoying coastal destinations and natural parks. They prefer direct bookings, which may indicate a tendency to plan their trips independently. Destinations targeting this group should emphasize tranquil, nature-based experiences, such as beach resorts or eco-friendly lodges.

4.1.6 Comparative Insights

Age and Travel Choices: Younger groups (Group 3) are more inclined toward adventure and urban exploration, while older groups (Group 2) lean towards cultural enrichment. Middle-aged groups (Groups 1, 4, and 5) exhibit diverse preferences, ranging from cities and mountains to nature and beaches.

Seasonal Preferences: Spring is the most favoured season across all groups, indicating its broad appeal for travel. Some groups, such as Groups 4 and 5, also extend their travel into summer, reflecting a preference for destinations that thrive in warmer weather.

Booking Methods: Direct booking is the most common method across all groups, highlighting the importance of clear and straightforward travel arrangements. Groups 1 and 4 also leverage online platforms, while Group 3 shows an affinity for travel agency services.

Destination Preferences: Cities and mountains are the most popular destinations, appealing to both younger and older travellers. Beach and nature destinations cater to Group 5, reflecting their unique interest in relaxation and outdoor activities. Cultural sites exclusively attract the oldest group (Group 2).

4.2 Regression Analysis

The study employed regression analysis to forecast and quantify one tourism parameter specifically the length of stay of 150 respondents.

4.2.1 Average Length of Stay

Prediction on tourists' average length of stay in their preferred destinations provides businesses and accommodation owners insights on their preparation procedures and resource allocation arrangements. Regression analysis was employed to predict individuals' duration of stay according to his age, gender, purpose of travel, travel season, and preferred destination. The parameters used in the regression analysis are indicated in table 4.

Table 4. Parameters and Results of Regression Analysis

Item	Parameters
Model	Linear Regression
predictors	age, gender, purpose of travel, travel season, and preferred destination
response	Average length of stay
Dataset size	150
Train-test ratio	70/30
Mean Square Error	0.0
R-squared	1.0
Regression Coefficients:	[3.41645462e-16 -1.53522533e-16 4.35097488e-15 -3.09773436e-15 1.00000000e+00]
Intercept	-3.552713678800501e-15

Table 4 reflects the parameters utilized to perform regression analysis. The plot between the actual data for the length of stay and the predicted data is shown in figure 8. For matched predictions, there is an overlapping visual for the actual and predicted values.

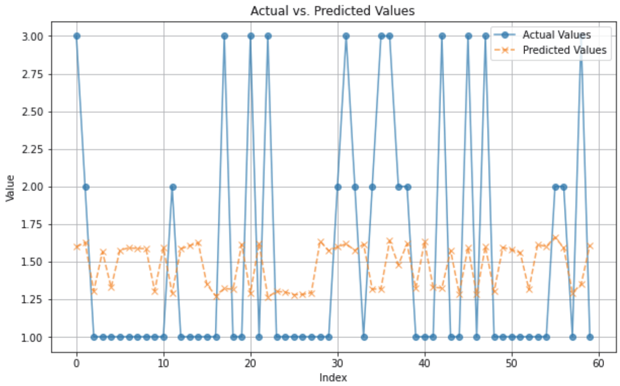

Figure 8. Actual vs Predicted data for Tourists' Average Length of Stay

Numerous tests were conducted to determine specific predictors significantly influence the tourists' average length of stay. Different combination of predictors was carried out where generated results are presented in table 5.

Legend:

A = gender, B = age, C = Purpose of Travel, D = Travel Season, E = Travel Destination

Table 5. Combination of Factors with Equivalent MSE

Predictors	MSE		Predictors	MSE
A,B	0.65	1	A,C,D	0.71
A,C	0.64	2	A,C,E	0.84
A,D	0.72	3	A,D,E	0.82
A,E	0.80	4	B,C,D	0.70
B,C	0.64	5	B,C,E	0.81
B,D	0.70	6	B,D,E	0.75
B,E	0.71	7	C,D,E	0.80
C,D	0.70	8	A,B,C,D	0.70
C,E	0.82	9	A,B,C,E	0.83
D,E	0.76	10	A,C,D,E	0.82
A,B,C	0.64	11	A,B,D,E	0.82
A,B,D	0.71	12	B,C,D,E	0.79
A,B,E	0.79	13	A,B,C,D,E	0.81

The above table reveals that the combination of three (3) predictors, A-Gender, B-Age and C-Purpose of Travel has generated the lowest MSE=0.64 indicating significant factors influencing tourists' length of stay in their preferred destinations.

5. Conclusion

This study provides a detailed exploration of traveller groups based on their average age, preferred travel seasons, destinations, and booking methods. By identifying unique characteristics and preferences, the findings offer valuable insights for tourism stakeholders seeking to enhance their services and attract a diverse range of travellers. The segmentation highlights the varying needs and priorities of different traveller demographics, emphasizing the importance of tailored approaches in the competitive tourism industry.

The preference for spring as a travel season across most groups reveals its broad appeal, likely due to its mild weather, natural beauty, and the availability of cultural events. However, the inclusion of summer by some groups, particularly those who enjoy nature and beach destinations, underscores the importance of catering to seasonal shifts in travel demand. These findings suggest that destinations should align their promotional strategies with seasonal trends to maximize visitor satisfaction and revenue.

In terms of destination preferences, cities and mountains emerge as popular choices for multiple groups, reflecting the allure of urban exploration combined with the tranquillity of nature. On the other hand, cultural sites and beaches cater to more niche audiences, such as older travellers seeking enrichment and middle-aged individuals prioritizing relaxation. This diversity highlights the importance of diversifying destination offerings to meet the varied expectations of travellers, whether they are seeking adventure, cultural immersion, or peaceful retreats.

The booking behaviour of these groups also provides actionable insights. The preference for direct booking across all groups signals the importance of having clear, user-friendly platforms that simplify the travel planning process. However, the reliance on online platforms and travel agencies by certain groups suggests that providing multiple booking options is critical for meeting travellers' diverse planning habits. Tourism operators can benefit from enhancing both direct and third-party booking channels to maximize their reach.

Understanding the age-related trends in travel behaviour is another significant takeaway from this study. Younger travellers tend to favour more adventurous and exploratory destinations, while older travellers prefer cultural and relaxing experiences. Middle-aged groups show a balanced approach, enjoying both natural and urban attractions. These insights underscore the value of designing age-specific travel packages that address the unique motivations and priorities of each demographic. Meanwhile, combination of gender, age and purpose of travel indicate significant predictors influencing tourists' duration of stay in their chosen destinations.

References

Aguilar, M. I., & Diaz, B. (2019). Length of stay of international tourists in Spain: A parametric survival analysis. *Annals of Tourism Research, 79*, 102768. https://doi.org/10.1016/j.annals.2019.102768

Bartl, E., Weigert, M., Bauer, A., Schmude, J., Karl, M., & Küchenhoff, H. (2025). Understanding travel behaviour patterns and their dynamics: Applying fuzzy clustering and age-period-cohort analysis on long-term data of German travellers. *European Journal of Tourism Research, 39*, 3914. https://doi.org/10.54055/ejtr.v39i.3862

Brida, J. G., & Scuderi, R. (2013). Determinants of tourist expenditure: A review of microeconometric models. *Tourism Management Perspectives, 6*, 28–40. https://dx.doi.org/10.2139/ssrn.2048221

Feng, L., Mu, L., Yang, Y., & Li, C. (2024). Research on tourist destination passenger flow prediction based on decomposition clustering technology for data continuity problem. In *2024 International Conference on Culture-Oriented Science & Technology (CoST)* (pp. 60–65). IEEE. https://doi.org/10.1109/CoST64302.2024.00021

Hasanah, H., Sudibyo, N. A., & Galih, R. M. (2021). Data mining using K-means clustering algorithm for grouping countries of origin of foreign tourist. *Nusantara Science and Technology Proceedings*, 88–94. https://doi.org/10.11594/nstp.2021.1112

Jun, W., Yuyan, L., Lingyu, T., & Peng, G. (2018). Modeling a combined forecast algorithm based on sequence patterns and near characteristics: An application for tourism demand forecasting. *Chaos, Solitons & Fractals, 108*, 136–147. https://doi.org/10.1016/j.glas.2018.01.028

Kaur, S., & Kaur, M. (2019). Behavioral intentions of heritage tourists: Influential variables on recommendations to visit. *Journal of Heritage Tourism, 15*(5), 511–532. https://doi.org/10.1080/1743873X.2019.1692852

Li, J., Weng, J., Shao, C., & Guo, H. (2016). Cluster-based logistic regression model for holiday travel mode choice. *Procedia Engineering, 137*, 729–737. https://doi.org/10.1016/j.proeng.2016.01.310

Li, K., Liang, C., Lu, W., Li, C., Zhao, S., & Wang, B. (2020). Forecasting of short-term daily tourist flow based on seasonal clustering method and PSO-LSSVM. *ISPRS International Journal of Geo-Information, 9*(11), 676. https://doi.org/10.3390/ijgi9110676

Monica, S., Natalia, F., & Sudirman, S. (2018). Clustering tourism object in Bali Province using K-means and X-means clustering algorithm. In *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)* (pp. 1462–1467). IEEE. https://doi.org/10.1109/HPCC/SmartCity/DSS.2018.00241

Peng, B., Song, H., Crouch, G. I., & Witt, S. F. (2015). A meta-analysis of international tourism demand elasticities. *Journal of Travel Research, 54*(5), 611–633. https://doi.org/10.1177/0047287514528283

Ruiz Reina, M. Á. (2021). Tourism and Big Data: Forecasting with hierarchical and sequential cluster analysis. *Engineering Proceedings, 5*(1), 14. https://doi.org/10.3390/engproc2021005014

Wang, S., Li, J., & Zhao, X. (2018). Exploring tourist behaviour with K-means clustering: Insights from geotagged social media. *Journal of Tourism Management, 70*, 1–15.

Yildirim, M. E., Kaya, M., & FurkanInce, I. (2022). A case study: Unsupervised approach for tourist profile analysis by K-means clustering in Turkey. *Journal of Internet Computing and Services, 23*(1), 11–17. https://doi.org/10.7472/JKSII.2022.23.1.11

Zhou, X., & Chen, Z. (2021). Destination attraction clustering: Segmenting tourist movement patterns with geotagged information. *Tourism Geographies, 25*(2–3), 797–819. https://doi.org/10.1080/14616688.2021.2006769