INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN SOCIAL SCIENCE (IJRISS)
ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume IX Issue VIII August 2025

Personalized Itinerary Generation Prototype Using K-Means Clustering and 2-Opt Optimization with Integrated User Content Sharing (in London)

Stephen Ebuka Iheagwara

Air Force Institute of Technology Kaduna, Kaduna, Nigeria

DOI: https://dx.doi.org/10.47772/IJRISS.2025.908000391

Received: 08 August 2025; Accepted: 14 August 2025; Published: 12 September 2025

ABSTRACT

Conventional travel planners often offer one-size-fits-all schedules and fixed tour packages, overlooking individual tastes and the wealth of peer insights. To enhance the personalization and interactivity of modern travel planning tools, this study presents a modular AI-based web application that generates optimized London itineraries tailored to individual preferences. The system integrates three core components. First, an AI-based itinerary generation module applies K-Means clustering—guided by the Elbow Method and Silhouette Analysis—to group Points of Interest (POIs) into thematic clusters that are both geographically compact and interest-aligned. Second, a travel route optimization module uses the 2-Opt algorithm, leveraging Haversine distance calculations to reorder each day's POIs, reducing intra-day travel distances by 12–18% on average. Third, a user-generated content module, built on Flask and SQLite, enables travelers to share blogs and vlogs and engage through likes and comments, fostering a socially enriched planning environment. The integration of algorithmic clustering and route optimization ensures that daily itineraries are spatially efficient and contextually relevant, while the social features heighten engagement and perceived value. Early user feedback indicates that the community-driven content enhances trust and decision-making quality. This dual-layered architecture bridges data-driven automation with collaborative user input, marking a shift from static itinerary tools to adaptive, socially-informed travel platforms. The proposed system offers a scalable framework for future smart tourism applications that aim to combine personalization, optimization, and participatory design.

INTRODUCTION

Travel planning has evolved from manual methods and static tools to intelligent, AI-powered systems capable of offering personalized experiences. Despite these technological advancements, many current platforms remain limited in functionality, particularly in terms of real-time adaptability and community engagement (Karthik & Vignesh, 2024). For instance, although London continues to attract a highly diverse tourist population, existing tools often fail to accommodate individual preferences or provide socially-informed recommendations (Rashid, 2023).

AI techniques, especially K-Means clustering, have significantly improved the grouping of Points of Interest (POIs) based on activity types and geographic location (Zubair et al., 2024). However, the integration of algorithmic itinerary planning with user-generated content remains largely underutilized (Mudhale et al., 2024). This is a critical shortcoming, given that peer-generated content—such as blogs and vlogs—has been shown to heavily influence travel decision-making (Jog & Alcasoas, 2023; S'hail & Benabdelouahed, 2024).

To bridge this gap, the present work proposes a modular, AI-enhanced travel planning system that combines automated itinerary generation with interactive user content sharing. The system applies K-Means clustering to form thematic, location-aware POI groupings aligned with user preferences and visualizes them through an intuitive web interface (Khatun, 2024; Veluru, 2023). In addition, it incorporates 2-Opt route optimization to minimize intra-day travel distances, resulting in more efficient and coherent itineraries.

Existing literature further emphasizes the limitations of traditional AI-based travel planners. Mariammal et al. (2022) observe that most systems rely on static datasets, lack adaptability, and fail to support interactive features such as experience sharing. While some systems adopt hybrid models to enhance itinerary generation, they often overlook the social dimension of real-world travel planning—where users share experiences, influence others, and contribute to a more participatory process. This underscores the need for a platform that is both algorithmically optimized and socially enriched.

This study addresses these challenges by developing a Flask-based web application composed of three integrated modules:

(1) **AI-Based Itinerary Generation Module** – Preprocesses and encodes a curated dataset of London POIs, then applies K-Means clustering using k (determined via the Elbow Method and Silhouette Analysis) to form thematic daily clusters. These clusters group POIs not only by spatial proximity but also by activity type. For example, one cluster may include the British Museum, the National Gallery, and the Tate Modern—forming a culture-focused itinerary, while another may feature Hyde Park, Kensington Gardens, and St. James's Park, reflecting a nature-oriented theme.

(2) **Travel Route Optimization Module** – Employs a 2-Opt heuristic, using Haversine distance, to optimize the sequence of POI visits within each cluster, reducing travel redundancy.

(3) **User-Generated Content Module** – Enables travelers to post blogs and embed vlogs, and to engage socially via likes and comments, thereby enriching the planning process through peer insights.

The primary aim of this project is to develop a smart itinerary planning prototype tailored to London that seamlessly integrates personalization, route optimization, and user participation. Specifically, the objectives are to:(i) implement K-Means clustering to generate interest-aligned itineraries, (ii) apply the 2-Opt algorithm to minimize intra-day travel distances, and (iii) develop a user-generated content module that fosters engagement and trust.

By combining machine learning techniques with participatory design, this system offers a dynamic, socially informed alternative to static itinerary tools—marking a step forward in the development of next-generation smart tourism platforms.

Building on the foundation laid by Mariammal et al. (2022), who proposed a hybrid machine learning approach for itinerary generation, this study extends their work by implementing a fully functional system that integrates clustering, route optimization, and user engagement. While their model employed K-Means and KNN for itinerary personalization, it lacked real-time interactivity and features for community-driven input. This paper addresses that gap by introducing a user-generated content module that allows travelers to upload blogs and vlogs, and interact through likes and comments—enriching the itinerary planning experience with peer insight and social validation. The system not only retains the technical benefits of machine learning–based itinerary generation and route efficiency but also enhances user engagement and trust through participatory content sharing. By unifying algorithmic planning with collaborative features in a working prototype, this work advances the development of adaptive, socially-informed smart tourism systems.

LITERATURE REVIEW

Recent advancements in artificial intelligence (AI) have significantly influenced the development of personalized travel tools. Mariammal et al. (2022) combined K-Means and KNN algorithms to generate itinerary recommendations based on user preferences and sentiment analysis. While technically effective, their system lacked real-time interactivity and social feedback mechanisms. Other studies have emphasized AI-driven automation (Priya, 2024), and adaptive tools like $TravelAgent$ (Chen et al., 2024), yet these platforms remain limited in collaborative engagement and dynamic user input. Beyond automation, effective travel personalization depends on how well systems can group destinations in a way that reflects user interests and spatial efficiency.

K-Means clustering has emerged as a popular approach for grouping POIs by location and interest. Zubair et al. (2024) and Xiao (2024) validated its suitability for spatial personalization, though both highlighted issues with centroid sensitivity and outlier handling—challenges addressed in this work through static datasets and preprocessing. Yee et al. (2024) applied K-Means in collaborative planning systems but noted challenges when user interests are highly diverse. Abhiram et al. (2024) further reinforced its use in organizing multi-day travel plans. While clustering establishes meaningful groupings of POIs, optimizing the travel route within those clusters is critical for minimizing effort and maximizing time efficiency.

Efficient route planning is critical to user satisfaction in travel systems. Jadhav et al. (2023) paired K-Means with the TSP for personalized and optimized itineraries. Gunawan and Iryanto (2023) improved route efficiency through a hybrid Simulated Annealing and 2-Opt model, while Sarawan and Khumla (2023) advanced this further with 2-OptRS using real map data. The current study adopts 2-Opt for its simplicity and proven effectiveness in minimizing intra-day travel distances. However, technical efficiency alone does not fully address the needs of modern travelers, whose choices are increasingly influenced by peer recommendations and shared experiences.

Social content plays an increasingly influential role in travel decisions. Jog and Alcasoas (2023) and S'hail and Benabdelouahed (2024) emphasized the persuasive impact of blogs and vlogs on destination choice. Despite this, most smart itinerary systems omit content-sharing features. Khatun (2024) applied NLP to unstructured social media data, showing promise in content analysis but highlighting the difficulty of working with noisy inputs. This system addresses the issue by structuring content inputs as embedded blogs and vlogs. Sai Mohith et al. (2025) and Chen et al. (2025) both called for more socially enriched and intuitive systems—precisely the dual focus of this project. Despite these valuable contributions, persistent limitations remain in system adaptability, interactivity, and modular scalability.

Although the body of literature on AI-driven itinerary planning continues to expand, most systems remain siloed in algorithmic design. Mariammal et al. (2022) identified the lack of interactive features as a key limitation. Others, including Mudhale et al. (2024) and Veluru (2023), highlighted issues with scalability, adaptability, and user feedback integration. Hermanto et al. (2024) stressed the importance of modularity for small-scale deployment—an approach adopted in this study using Flask. This project builds upon prior strengths while directly addressing gaps in community integration, personalization, and real-time user participation.

METHODOLOGY

This section describes the architecture, modules, data preparation, clustering, route optimization, and content-sharing mechanisms that underpin the system. The goal was to develop a personalized and optimized trip planning tool that combines machine learning techniques, geographic algorithms, and user-generated content for enhanced user engagement and route efficiency.

System Architecture and Module Overview

The system is built using a modular architecture that integrates three major components:

- AI-Based Itinerary Generation Module
- Travel Route Optimization Module (2-Opt)
- Content Sharing Module

The interaction among these components is coordinated through a Flask-based backend connected to an SQLite database, and rendered via an HTML/CSS/JavaScript frontend.

Dataset Used

The dataset used for this project was obtained from Mendeley Data. It contains information about Points of Interest (POIs) in London, including categories such as nature, nightlife, drink, music, sports, art, museum, and more. It was preprocessed by replacing missing values with 0 to ensure consistency for machine learning applications such as K-Means clustering. Additionally, the dataset was enhanced by obtaining precise geographical coordinates for each POI using the OpenStreetMap API, specifically leveraging the Nominatim service for geocoding.

AI-Based Itinerary Generation

This module is responsible for recommending day-by-day POI groupings based on user-selected activities and travel duration. It employs a K-Means clustering approach that integrates two primary input features:

- Binary vectors representing selected activity categories (user interests)
- Geographical coordinates (latitude and longitude) of POIs

This ensures that the resulting clusters are both geographically compact, minimizing travel time between POIs, and semantically relevant, aligning with the user's selected interests.

Determining Optimal k Two methods were used to determine the optimal number of clusters (k):

- *Elbow Method*: For k = 2 to 20, the within-cluster sum of squares (WCSS) was computed. A clear elbow was observed at k = 10, where further clusters showed diminishing WCSS reduction (sharp WCSS drop from ~2,000 to ~1,000 at k = 10, tapering to ~500 by k = 20).

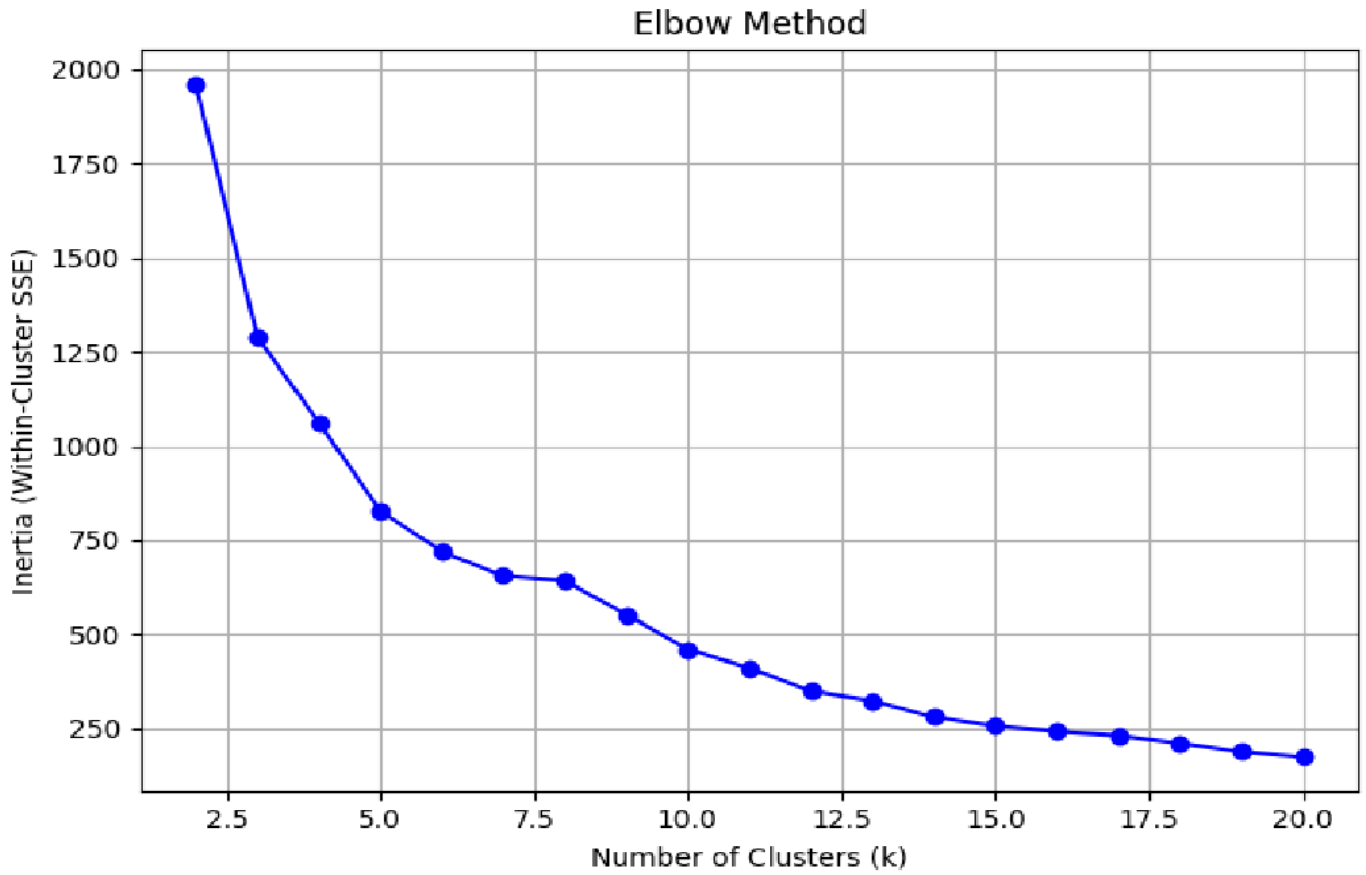

Figure 1: Elbow method to obtain optimal k

- *Silhouette Analysis*: For each k in the same range, the mean silhouette score was calculated. Mean silhouette scores rose from 0.40 at k = 2 to 0.68 at k = 10, peaking at 0.85 by k = 20. However, the improvement beyond k = 10 was marginal, indicating that 10 clusters strike the best balance of cohesion and separation.

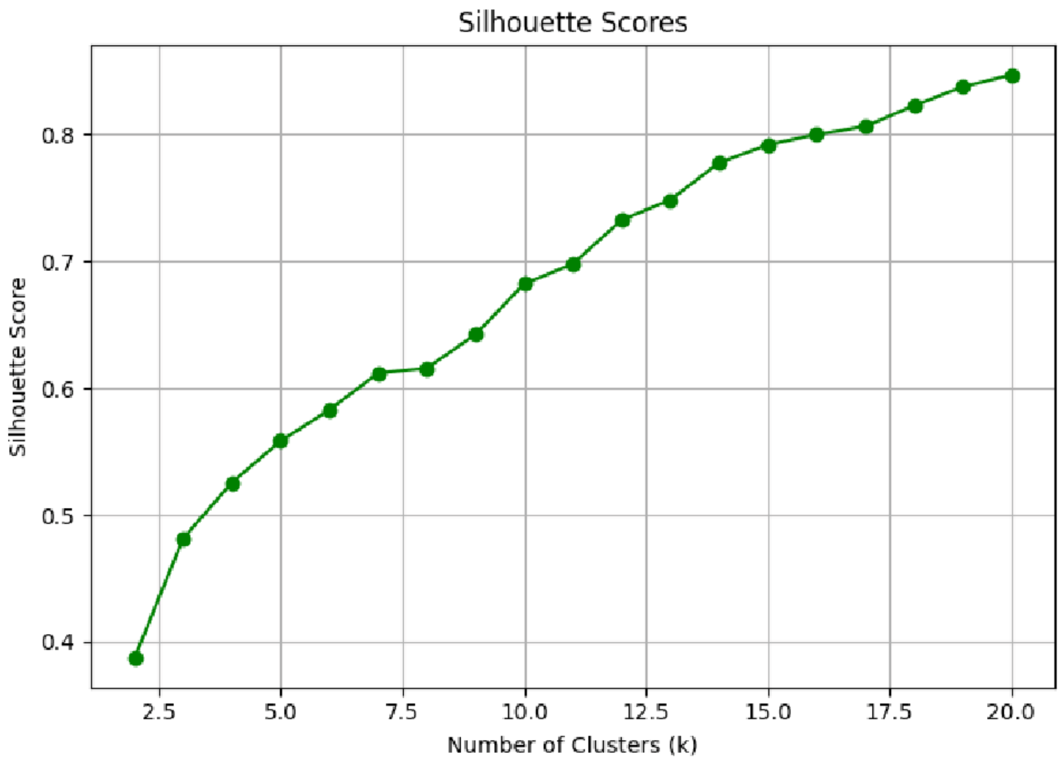

Figure 2: Silhouette score graph

Clustering and Itinerary Construction: With k = 10, K-Means assigned each POI to one of ten clusters. For a u-day trip, clusters were ranked by proximity to cluster centroids and matched to travel days proportionally. Within each assigned cluster/day, the N closest POIs to the centroid were selected to ensure thematic relevance. Final itineraries are presented as day-wise lists of POIs, annotated with matched activities and coordinates.

Pseudocode for Itinerary Generation

Load and Preprocess POIs

- LOAD(dataset): Read POI dataset from CSV file
- REMOVE_NULL_COORDINATES(dataset): Eliminate entries missing latitude or longitude
- NORMALIZE(dataset): Apply Min-Max scaling to geographic and activity features to ensure equal weighting in clustering

Filter POIs by Selected Activities

- FILTER(dataset, selected_activities): Return only POIs that match any selected activity category
  Cluster POIs Using K-Means
- DEFINE_FEATURES: Use normalized latitude, longitude, and selected activity columns
- APPLY_KMEANS(features, k): Perform K-Means clustering with a fixed number of clusters k
- ASSIGN_CLUSTER_LABELS(POIs, kmeans_model): Assign each POI to its corresponding cluster
- EVALUATE_CLUSTERING(kmeans_model): Compute silhouette score to assess clustering quality

Select Representative POIs from Each Cluster

- For each cluster:
  - COMPUTE_DISTANCES(cluster_center, POIs): Calculate Euclidean distance from center to POIs
  - SELECT_TOP_N_POIs(cluster, N): Choose the N closest POIs to the cluster center

Build Daily Itineraries

- REMOVE_DUPLICATES(POIs): Remove duplicate POIs based on name
- SORT_BY_RELEVANCE(POIs, selected_activities): Rank POIs by how well they match selected activities
- SELECT_TOP_POIs = days × N: Choose top POIs needed for all days
- SPLIT_INTO_DAYS(POIs, days): Evenly divide POIs into days sublists
- For each day:
  - ASSIGN_POIs(day, POI_list): Include POI name, coordinates, and matched activities
- RETURN(itineraries): Output the list of daily POI assignments

Algorithm 1: Pseudocode for the itinerary generation pipeline using K-Means clustering on geographic and activity-based features.

```
Silhouette Score for k=10: 0.9619
Day 1: 5 POIs selected.
Day 2: 5 POIs selected.
Day 3: 5 POIs selected.
Day 4: 5 POIs selected.

Itinerary for Day 1:
 - The Sun (North street)
 - Party Knightz (Mulberry Parade, West Drayton, United Kingdom)
 - The White House - Stockley Park (Stockley Park, London, UB11 1AA, United Kingdom)
 - Bacchus Late Bar and Venue (2 Union Street, Kingston upon Thames, United Kingdom)
 - Oceana (154 Clarence Street, Kingston upon Thames, United Kingdom)

Itinerary for Day 2:
 - The Corrib Rest (82 Salisbury Rd)
 - The Sussex Arms (5-27 Staines Rd)
 - Brouge at the Old Goat (241 Hampton Road, Twickenham, TW2 5NG, United Kingdom)
 - The Rifleman Pub Twickenham (7 Fourth cross road, Twickenham, TW2 5EL, United Kingdom)
 - Honeycombe Hungryhorse (417 STAINES ROAD, Hounslow, TW4 5AR, United Kingdom)

Itinerary for Day 3:
 - Strawberry Hill House (268 Waldegrave Road, Twickenham, United Kingdom)
 - Eel Pie Pictures (Eel Pie Island, Twickenham, United Kingdom)
 - Greenwich Heritage Centre (Artillery Square, Royal Arsenal, London, United Kingdom)
 - Carlyle's House (24 Cheyne Row, Chelsea, United Kingdom)
 - Thames Barrier (1 Unity Way)

Itinerary for Day 4:
 - Alexandra Palace (Alexandra Palace Way, London, United Kingdom)
 - Finsbury Park (Endymion Road, London, United Kingdom)
 - Potters Fields Park (London, United Kingdom)
 - Earls Court Exhibition Centre (Warwick Road, London, United Kingdom)
 - Syon Park (Syon Park, London Road, Brentford, United Kingdom)
```

Figure 3: 4 days Itinerary Generated from Selected Activities

('nightlife','dance','music','walk','history') and Geographic Clustering Using K-Means Algorithm

Travel Route Optimization (2-Opt)

Although clustering ensures logical groupings of POIs per day, the sequence in which they are visited is not initially optimized. To address this, the system uses the 2-Opt algorithm, a local search heuristic for the Travelling Salesman Problem (TSP). This optimization significantly improves route efficiency within each day, reducing travel time and enhancing user experience.

To minimize intra-day travel distances, this module applies the 2-Opt algorithm, optimizing the sequence in which POIs within a daily cluster are visited. Haversine distance is used as the distance metrics.

- *Initial Route*: POIs for a given day were initially ordered by selection ranking.
- *2-Opt Iterations*: All non-adjacent edge pairs (i, j) were considered. For each pair, the sub-sequence between i and j was reversed, and the total travel distance (via Haversine formula) was re-evaluated. If the new sequence reduced distance, it was accepted. This process continued until no further improvements were found.

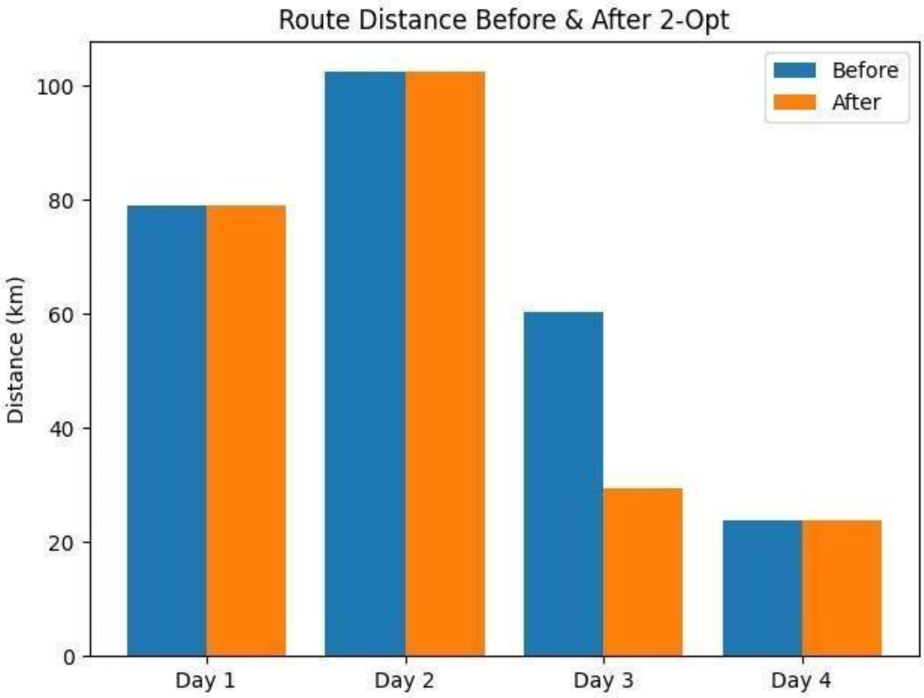

Figure 4: Distance reduction achieved by 2-Opt for each travel day.

Distance Metric: Haversine Formula To compute the great-circle distance between two POIs:

$$d = 2r \times \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\emptyset}{2}\right) + \cos(\emptyset_1) \times \cos(\emptyset_2) \times \sin^2\left(\frac{\Delta\lambda}{2}\right)}\right)$$

Where:

- $\emptyset_1,\ \emptyset_2 =$ latitudes (in radians)
- $\lambda_1,\ \lambda_2 =$ longitudes (in radians)
- $\Delta\emptyset = \emptyset_2 - \emptyset_1$
- $\Delta\lambda = \lambda_2 - \lambda_1$
- $r =$ Earth's radius (typically **6371 km**)

This provides accurate real-world travel estimates.

Content Sharing Module

This module allows users to upload personal travel blogs and vlogs. Other users can interact with the content through likes and comments, promoting a community-based planning approach.

Upload Capabilities:

- Travel Blogs: Rich text-based trip descriptions
- Travel Vlogs: Embedded video links (e.g., YouTube)

Interaction Features:

- Users can like and comment on posts
- Popular uploads are highlighted to promote community trust and discoverability
- Each post is associated with a user and timestamp

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume IX Issue VIII August 2025

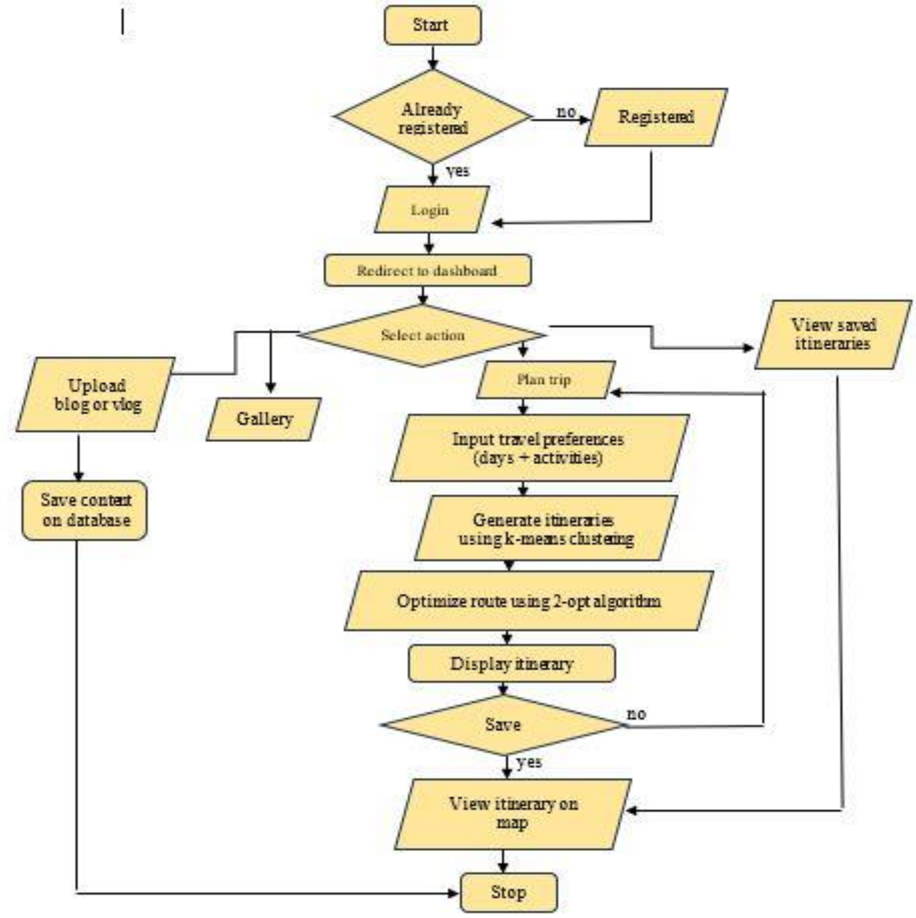

Figure 5: Flowchart of the system

Database Schema

User Table: Stores user data such as ID (primary key), username, email, password (hashed), date joined, and admin status.

	id	username	email	password_hash	date_joined	is_admin
1	1	victoria	rvictoria786@gmail.com	$2b$12$HBR7ziWeaqPh1ZJp3crg.rgpFrtbyXGTLNzfqCV...	2025-04-09 17:28:59.738913	0
2	2	victoria1	vnn@gmail.com	$2b$12$.UNG79KmVXmtFQxOudTKleoHr8JL892kDOEw...	2025-04-10 18:17:41.808219	0
3	3	admin	admin@example.com	$2b$12$nK0Gkgjhc1T/QARG1qiPlO6i1.9PyCcIS.2bv.qR/...	2025-04-11 22:59:44.151768	1
4	5	Aisha	aishaolamide13@gmail.com	$2b$12$niv/pNs1DgshQhnPG/OZcORUKgjZLGWXi/Leu0...	2025-04-12 11:20:54.301007	0
5	6	Kayla	kayla001@gmail.com	$2b$12$t/rsX.WUU/7jr6LKO0gAu8W3olfkJjSc0XpNOiwl...	2025-04-12 11:52:28.442601	0
6	7	Rhodiyah	rhodiyaho@gmail.com	$2b$12$xbrANM8I9/DMZ8FQ8ni2D.SMbHDzI0Y4j3O3Iq...	2025-04-12 11:59:18.755697	0
7						

Figure 6: User Table

Upload Table: Stores blog text or vlog URL/file path, timestamp, and uploader reference.

	id	user_id	upload_t...	filename	vlog_url	timestamp	vlog_title
1	1	1	blog	blog1.txt	NULL	2025-04-11 21:40:15	NULL
2	2	1	vlog	NULL	https://www.youtube.com/watch?v=gT3ebVxcozw	2025-04-11 21:40:36	NULL
3	3	2	vlog	NULL	https://www.youtube.com/watch?v=WJxRRZ24M-4	2025-04-11 22:21:29	NULL
4	4	5	blog	blog3.txt	NULL	2025-04-12 11:28:28	NULL
5	5	6	vlog	NULL	https://www.youtube.com/watch?v=uwVl6zCzctl	2025-04-12 11:56:13	london vlog
6	6	7	blog	my_experience.txt	NULL	2025-04-12 12:06:50	NULL
7	7	1	blog	blog1.txt	NULL	2025-04-12 14:25:53	NULL
8	8	1	vlog	NULL	/static/uploads/3007578-uhd_3840_2160_30fps.mp4	2025-05-08 16:53:23	My View
9							

Figure 7: Upload Table

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume IX Issue VIII August 2025

Comment Table: Links user comments to uploads.

	id	user_id	uploa...	content	timestamp
1	1	2	1	great	2025-04-11 22:22:28.350641
2	2	1	2	interesting	2025-05-11 15:54:29.141585
3					

Figure 8: Comment Table

Like Table: Records user “likes” per upload, preventing duplicates via unique constraints.

	id	user_id	uploa...
1	1	2	1
2	2	5	3
3	3	7	5
4	4	1	4
5	5	1	5
6	6	1	6
7	7	1	1
8			

Figure 9: Like Table

Backend Routes (Flask)

- /upload (GET/POST): Renders upload form; on POST, saves text or file and creates Upload record.
- /gallery (GET): Fetches uploads, comments, and like counts; renders via Jinja2 templates.
- /like_upload/< id > (POST): Toggles like status for the current user.
- /comment_upload/< id > (POST): Appends a comment to the specified upload.

Frontend Interface

Responsive pages were built with Bootstrap 5, displaying uploads in card grids. Blogs show titles and excerpts; vlogs are embedded via <iframe> for YouTube links. Interactive “Like” buttons and comment forms accompany each upload.

Module Evaluation

Functional testing confirmed successful upload, retrieval, and display of both blogs and vlogs. Users can like and comment without duplication. UI feedback (e.g., updated like counts) occurs in real time.

Technology Stack

The system was implemented as a full-stack web application using the following technologies:

- **Frontend**: HTML, CSS, JavaScript, and Bootstrap 5 for responsive design
- **Backend**: Python (Flask framework)
- **Database**: SQLite for lightweight data storage
- **Geocoding Service**: OpenStreetMap Nominatim API (via the Geopy library)
- **Machine Learning**: Scikit-learn for K-Means clustering

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume IX Issue VIII August 2025

- **Optimization Algorithm**: Custom implementation of the 2-Opt heuristic using the Haversine formula
- **Templating Engine**: Jinja2 for rendering dynamic frontend content

RESULTS AND DISCUSSION

This section presents the user interface and observed performance of the proposed system. It includes a walkthrough of system navigation, evaluation of clustering and optimization modules, insights into the content-sharing component, and highlights of limitations encountered.

User Interface and System Navigation

The system provides a user-friendly interface with intuitive navigation paths for all users. Core pages include the Home page, Registration and Login forms, the Dashboard, and dedicated itinerary planning pages. Users can seamlessly move from selecting activities to generating itineraries and visualizing routes on an interactive map. These screens collectively support itinerary generation, visualization, and community interaction within a single environment.

Figure 10: Home Page

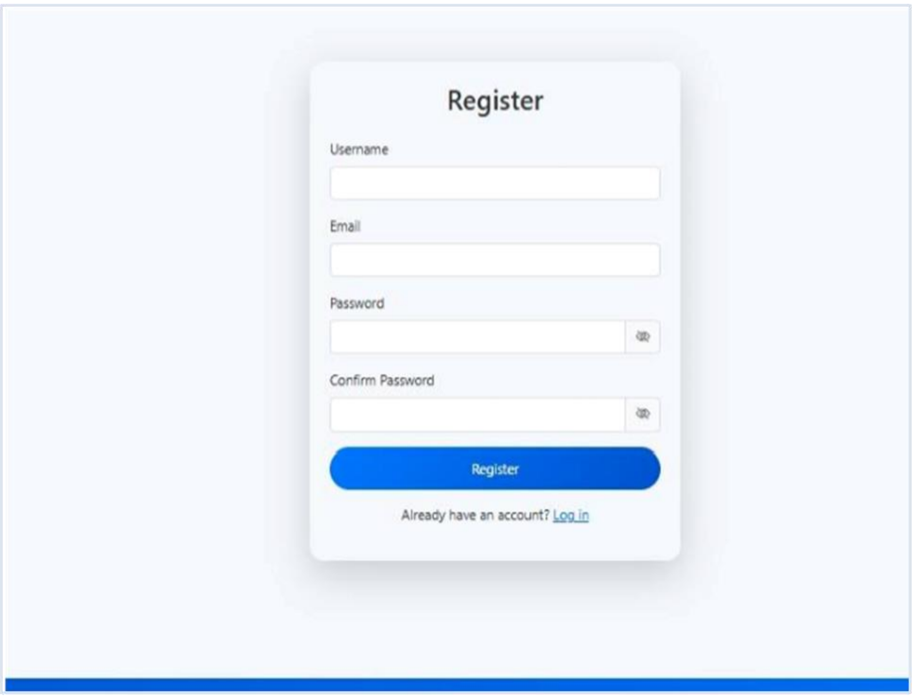

Figure 11: Register page

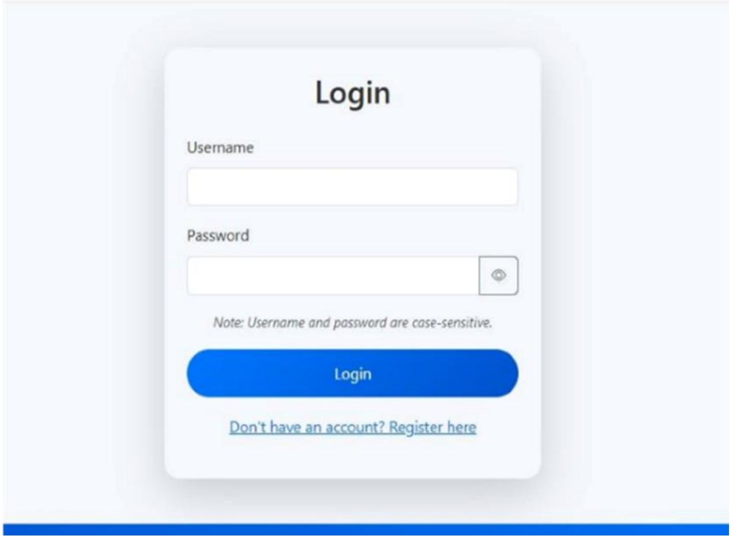

Figure 12: Login Page

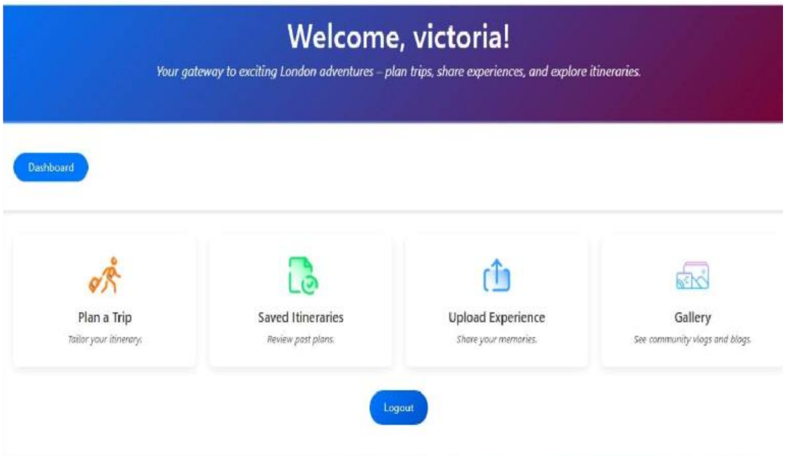

Figure 13: Dashboard

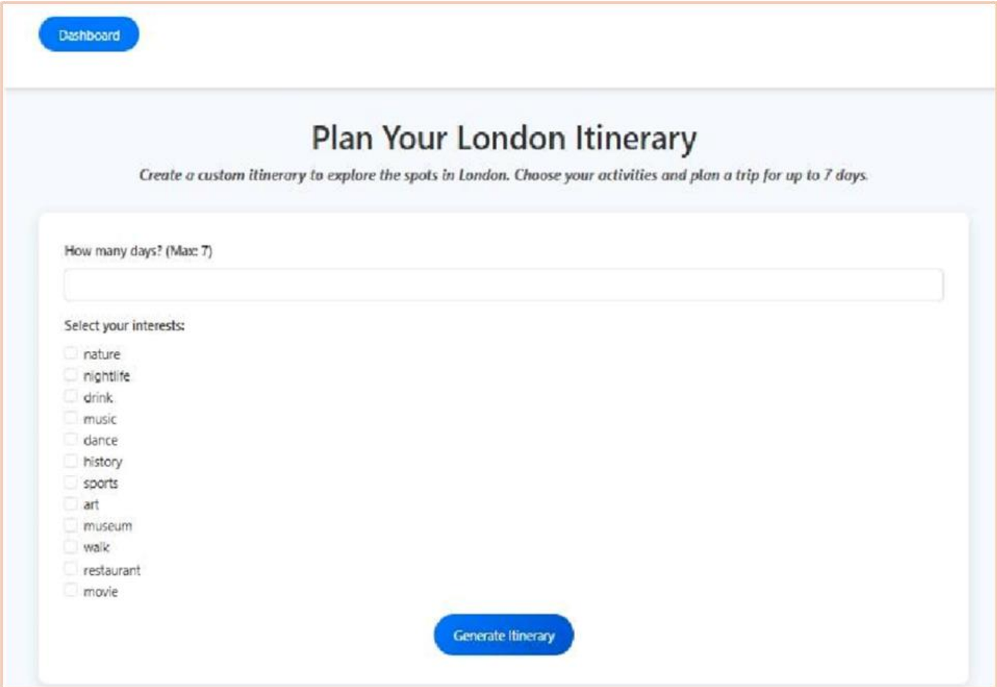

Figure 14: Plan itinerary page

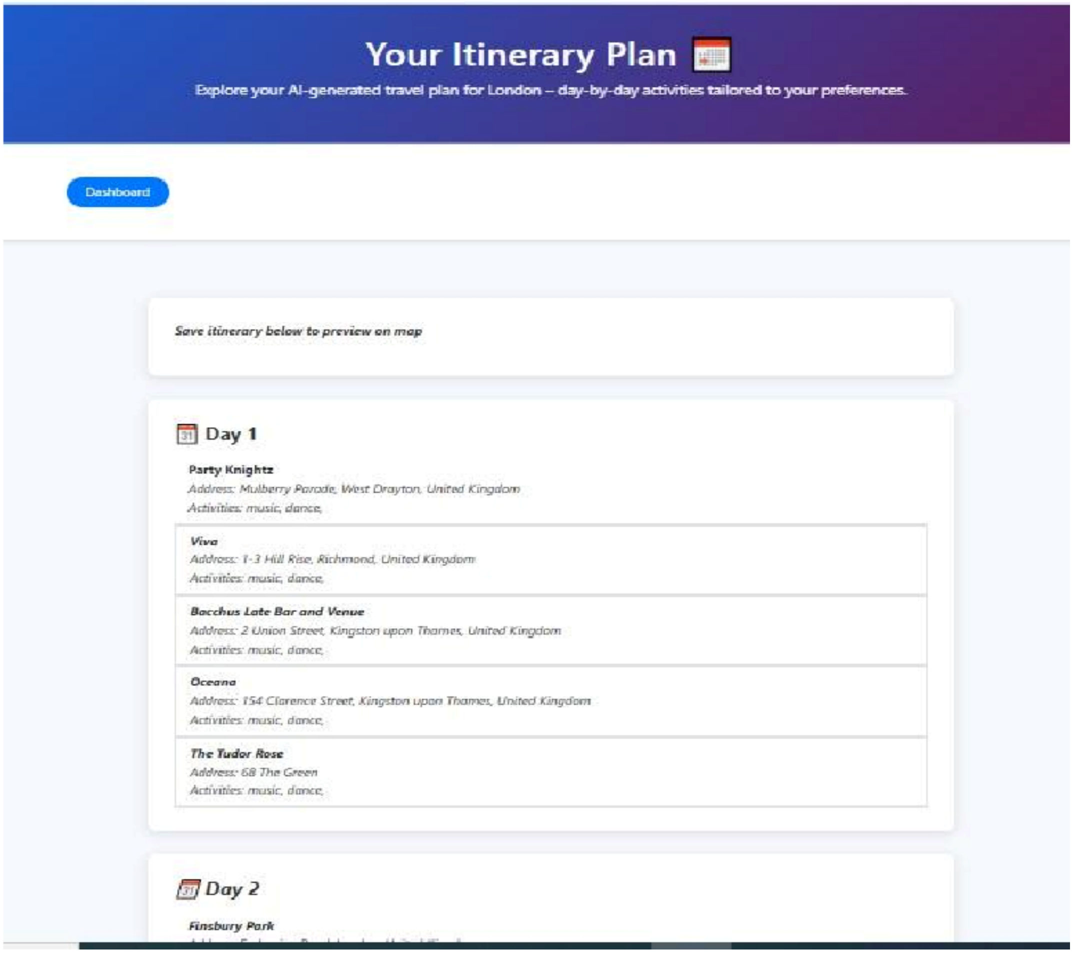

Figure 15: Generate Itinerary page

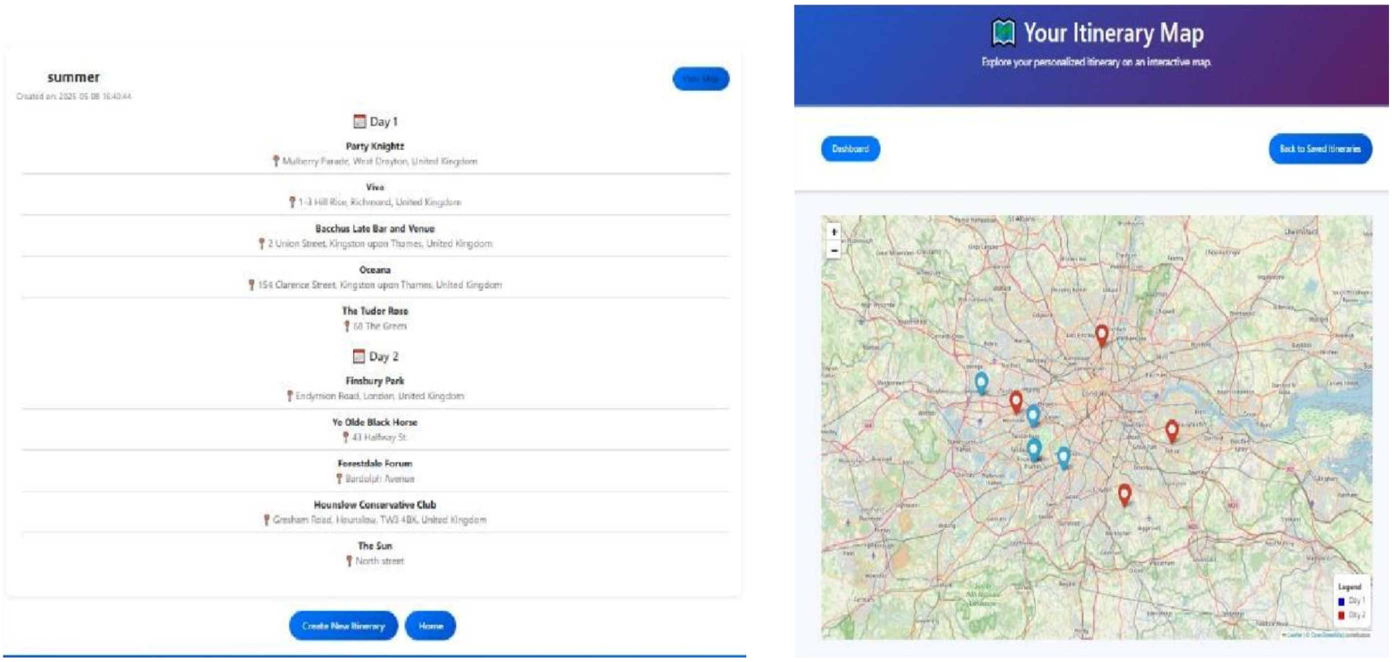

Figure 16: Itinerary Map page

Clustering Performance and Itinerary Structuring

A core element of the system involved K-Means clustering for grouping Points of Interest (POIs) into daily itineraries. By combining geographic coordinates with user-selected activity categories, the clustering ensured that each day's POIs were both geographically proximate and thematically aligned with user preferences.

Using the Elbow Method and Silhouette Score, the optimal cluster number was found to be **k = 10**, achieving a silhouette score of approximately **0.68**. Although silhouette values continued to rise for higher k, the improvement diminished beyond 10, confirming this choice as a balance between intra-cluster cohesion and inter-cluster separation. This resulted in logically structured and interest-driven itineraries.

Route Optimization and Travel Efficiency

While clustering created meaningful groups of POIs, the sequence of visits within each day required optimization. The 2-Opt algorithm was applied to minimize intra-day travel distances, with the Haversine formula ensuring accurate great-circle distance calculations. This optimization consistently reduced travel time and produced smoother travel routes, enhancing usability and real-world feasibility of the generated itineraries.

Content Sharing and Community Value

Beyond itinerary generation, the system incorporated a user-generated content (UGC) module. Users could upload blogs and vlogs, interact through likes and comments, and discover trending contributions. This layer transformed the system from a purely algorithmic planner into a **community-oriented platform** that fosters trust, inspiration, and peer-to-peer engagement. By allowing travelers to draw insights from shared experiences, the module enriched decision support and provided social credibility to itinerary suggestions. However, the current implementation lacks automated moderation and popularity-based filtering, which are necessary to ensure quality, relevance, and safety of shared content.

System Evaluation

The evaluation of the system was primarily based on clustering performance (silhouette score) and route efficiency (distance reduction after 2-Opt optimization). While these technical metrics validated the system's internal logic, the absence of usability testing or comparative benchmarking remains a limitation. Future work should include user surveys, performance testing under larger datasets, and benchmarking against existing itinerary planners to strengthen empirical validation

Limitations

Despite its strengths, the system has notable limitations. It is geographically restricted to London and relies on a static POI dataset without real-time updates on schedules, events, or traffic. Practical constraints such as weather, fatigue, or accessibility were not modeled, reducing real-world feasibility. The UGC module currently depends on manual moderation without automated tools for quality control. Finally, the system lacks mobile and offline support, which restricts accessibility in low-connectivity contexts.

CONCLUSION AND FUTURE WORK

This study addressed the challenges of traditional travel itinerary planning by developing an AI-driven system that integrates K-Means clustering for interest-based grouping, 2-Opt optimization for efficient routing, and a novel user-generated content module. By introducing a community-oriented platform where travelers can share blogs and vlogs, the system extends prior models and enhances trust, inspiration, and peer-to-peer engagement. Designed specifically for London, the platform demonstrated the ability to generate structured, personalized itineraries while enriching decision support with social content.

To advance the prototype into a fully deployable smart tourism platform, several enhancements are proposed. First, geographic scope will be expanded beyond London by integrating larger POI datasets through APIs such as Google Places, OpenStreetMap, TripAdvisor, and Foursquare. Real-time integration will allow dynamic updates on POI status, operating hours, live events, weather, and traffic. To handle this broader and dynamic dataset, the database will be migrated from SQLite to scalable geospatial storage solutions such as PostgreSQL/PostGIS or MongoDB.

Second, itinerary feasibility will be strengthened by incorporating practical travel constraints. This includes modeling POI opening hours as time windows, extending the clustering and routing algorithms into a time-constrained Travelling Salesman Problem (TSP), and adding daily time budgets to account for user fatigue. Weather adaptation will be supported by tagging POIs as indoor or outdoor and dynamically re-ranking recommendations using APIs such as OpenWeatherMap.

Third, system evaluation will go beyond technical clustering metrics by including usability studies with real users (e.g., surveys and System Usability Scale scores), comparative benchmarking with existing itinerary planners (e.g., Google Trips, Sygic, TripIt), and performance testing under larger datasets and higher concurrency. These evaluations will provide stronger empirical evidence of usability, responsiveness, and scalability.

Fourth, the user-generated content module will be reinforced with automated moderation and personalization. Natural language processing models such as spaCy, fastText, or HuggingFace will filter offensive or irrelevant text, while lightweight computer vision tools will detect inappropriate images or videos. Popularity-aware re-ranking using engagement metrics (likes, views, comments) will surface high-quality contributions, and hybrid recommender systems combining content-based and collaborative filtering will provide more personalized content suggestions.

Finally, deployment will extend to mobile platforms through Progressive Web Apps (PWAs) or native Android/iOS applications, with offline functionality supported by local caching and synchronization. This will ensure accessibility in low-connectivity environments and improve adoption by travelers on the move.

By following this roadmap, the system can evolve from a London-specific prototype into a scalable, adaptive, and globally relevant smart tourism platform, balancing technical efficiency with user-centered trust, transparency, and engagement.

REFERENCES

1. Abhiram H. R., Aravind S. K., Bharath R. Sindhe, Charan G. S., & Vanishri Sataraddi. (2024). Exploring automated systems for planning travel itineraries. *International Journal of Innovative Science and Research Technology, 9*(11), 3188-3195. https://doi.org/10.5281/zenodo.14545931
2. Chandra Sekhar Veluru. (2023). Transforming travel planning: The impact of generative AI on itinerary optimization, cost efficiency, and user experience. *Online Scientific Research*. Retrieved from https://www.onlinescientificresearch.com/articles/transforming-travel-planning-the-impac t-of-generative-ai-on-itinerary-optimization-cost-efficiency-and-user-experience.html
3. Chen, A., Ge, X., Fu, Z., Xiao, Y., & Chen, J. (2024). TravelAgent: An AI assistant for personalized travel planning. *arXiv Preprint*. https://doi.org/10.48550/arXiv.2409.08069
4. Chen, B., & Liu, J. (2025). User experience improvement through personalized travel planning: A study of AI travel customization. *Applied and Computational Engineering, 131*, 142–148. https://doi.org/10.54254/2755-2721/2024.20575 .
5. Gunawan, P. H., & Iryanto, I. (2023). Simulated annealing – 2 opt algorithm for solving traveling salesman problem. *International Journal of Computing*, *22*(1), 82–88. https://www.computingonline.net/computing/article/view/2878.
6. Hermanto, H. (2024). Implementation of the web-based K-means clustering algorithm on hypertension levels in the elderly at the Bungah District Health Center. *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics, 6*(2). https://doi.org/10.35882/h6596074
7. Jadhav, S., Kothari, D., Gori, D., Joshi, D., & Mehta, B. (2023). PlanIT: An itinerary planner using K-means clustering and Travelling Salesman Problem. *Proceedings of ICCUBEA 2023*, 1–4. https://doi.org/10.1109/ICCUBEA58933.2023.10392185
8. Jog, D., & Alcasoas, N. (2023). Travel decision-making through blogs and vlogs: An empirical investigation on how user-generated content influences destination image. *Turyzm/Tourism, 33*, 19–28. https://doi.org/10.18778/0867-5856.33.2.02
9. Karthik, B., & Vignesh, S. (2024). Intelligent travel planning insights using machine learning.

International Journal of Creative Research Thoughts. Retrieved from https://ijcrt.org/papers/IJCRT2404077.pdf

10. Khatun, T. (2024). AI-driven personalized travel recommendations based on real-time social media analysis. *ResearchGate*. https://www.researchgate.net/publication/383137002_AI-Driven_Personalized_Travel_R ecommendations_Based_on_Real-Time_Social_Media_Analysis.
11. Mohith, S. S., Hemamalini, H., Karthik, K., Nithin, N., & Sanath, R. (2025). A review paper on AI-driven travel planning. *Zenodo*. https://doi.org/10.5281/zenodo.14591573
12. Mudhale, A., Shirmale, M., Kudalkar, V., Motiray, R., & Ahmad, S. (2024). Travel itinerary planner using AI. *International Research Journal of Engineering and Technology (IRJET), 11*(4). https://www.irjet.net/archives/V11/i4/IRJET-V11I4152.pdf.
13. Padma Priya, A. (2024). An automated itinerary planning and trip management system. *International Journal of Creative Research Thoughts*. Retrieved from https://www.ijcrt.org/papers/IJCRT24A5217.pdf.
14. Sangeetha Mariammal, S. S., Akshaya, S. B., Priyanga, M., Saran Kumar, S., & Prakash, P. (2022). Smart travel assistant with itinerary planner using hybrid machine learning approach. *International Research Journal of Modern Engineering and Technology, 4*(5). Retrieved from https://www.irjmets.com/uploadedfiles/paper/issue_5_may_2022/22455/final/fin_irjmets1652185712 .pdf
15. Sarawan, K., & Khumla, P. (2023). Efficient route planning in Northeast Thailand using a random shuffle enhanced 2-Opt algorithm. *Journal of Information Systems and Technology Research*, *5*(2), 65–78. https://journal.aira.or.id/index.php/jistr/article/view/1017
16. S'hail, M., & Benabdelouahed, R. (2024). Exploring user-generated content in tourism: Insights from Moroccan travelers. *Tourism Studies Journal, 2*. https://doi.org/10.5281/zenodo.11045564.
17. Rashid, S. S. (2023, March 22). Potential of tourism sector in London and way forward – OpEd. *Eurasia Review.* https://www.eurasiareview.com/22032023-potential-of-tourism-sector-in-london-and-way-forward-oped/
18. Xiao, E. (2024). Comprehensive K-means clustering. *Journal of Computer and Communications, 12*, 146–159. https://doi.org/10.4236/jcc.2024.123009 .
19. Yee, L. Y., Belaidan, S. L. M., Rahman, N. A. A., & Harun, K. S. (2024). Implementing K-means clustering algorithm in collaborative trip advisory and planning system. *Academia Preprint*. https://www.academia.edu/114579766/Implementing_k_means_clustering_algorithm_in_ collaborative_trip_advisory_and_planning_system.
20. Zubair, M., Iqbal, M. A., Shil, A., & Others. (2024). An improved K-means clustering algorithm towards an efficient data-driven modeling. *Annals of Data Science, 11*(6), 1525–1544. https://doi.org/10.1007/s40745-022-00428-2.