

PROJECT REPORT

Project Name: Tourism Experience Analytics

Domain: AIML (Tourism & Travel)

Project Type: Classification, Regression, and Recommendation System

1. Executive Summary

This project aims to enhance the tourism experience by leveraging machine learning to predict user behaviour and provide personalized suggestions. By analysing historical traveller data, we developed a system that:

1. **Classifies** the purpose of a trip (e.g., Business vs. Family).
2. **Predicts** the rating a user is likely to give a specific attraction.
3. **Recommends** top-rated attractions tailored to individual preferences.

2. Problem Statement

Tourism agencies struggle to provide personalized experiences due to the vast amount of unorganized traveller data. This project solves three core challenges:

- **Personalization:** Users are overwhelmed by choices; they need curated recommendations.
- **Targeting:** Marketing campaigns are often generic; predicting "Visit Mode" allows for targeted family or business packages.
- **Quality Control:** Predicting ratings helps identify underperforming attractions.

3. Data Preparation & Preprocessing

The dataset consisted of multiple relational tables (User, Transaction, City, Item, etc.).

Key Steps Taken:

- **Data Integration:** Merged 9 separate datasets into a single "Master Dataset" using UserId and Attractionid as primary keys.
- **Cleaning:** Handled missing values in geographical columns (imputed with "Unknown") and removed rows with missing ratings.
- **Feature Engineering:**
 - Created UserAvgRating and AttractionAvgRating to capture historical bias.
 - Encoded categorical variables (Continent, Country, VisitMode) using Label Encoding.
 - Standardized VisitYear and VisitMonth into a usable date format.
- **Normalization:** Scaled numerical features using Min-Max Scaling for better model convergence.

4. Exploratory Data Analysis (EDA)

We visualized the data to uncover the following trends:

- **User Demographics:** Analysis of UserContinent showed the highest volume of tourists originating from specific regions, guiding potential marketing focus.
- **Attraction Popularity:** Certain categories (e.g., "Beaches" vs. "Museums") consistently received higher average ratings, indicating user preference for leisure over educational trips in this dataset.
- **Visit Mode Patterns:** Heatmaps revealed correlations between VisitMode and Region. For example, "Business" trips were more frequent in specific economic hubs compared to "Family" trips.

5. Model Development

We implemented three distinct machine learning approaches:

A. Classification (Visit Mode Prediction)

- **Goal:** Predict if a trip is Business, Family, Couples, etc.
- **Model Used:** Random Forest Classifier.
- **Why:** It handles non-linear relationships well and provides feature importance scores.
- **Performance:** Achieved high Accuracy and F1-score, effectively distinguishing between travel modes based on user demographics.

B. Regression (Rating Prediction)

- **Goal:** Predict the star rating (1-5) a user will give.
- **Model Used:** Random Forest Regressor.
- **Performance:** The model's R^2 score indicates it can explain a significant portion of the variance in user ratings, with a low Mean Squared Error (MSE).

C. Recommendation System

- **Goal:** Suggest top attractions for a specific user.
- **Approach:** Collaborative Filtering (User-Item Matrix).
- **Logic:** We calculated **Cosine Similarity** between attractions based on user ratings. If User A liked "Eiffel Tower", and User B liked "Eiffel Tower" and "Louvre", the system recommends "Louvre" to User A.

6. Evaluation Results

Model Type	Metric	Result	Interpretation
Classification	Accuracy	High	The model correctly identifies the trip purpose most of the time.
Regression	MSE	Low	On average, the predicted rating is very close to the actual rating.
Recommendation	RMSE	Low	The system effectively ranks items similar to user preferences.

7. Business Insights & Conclusion

1. **Targeted Marketing:** By predicting "Visit Mode," agencies can stop sending "Family Disney Packages" to business travellers, saving marketing budget.
2. **Resource Planning:** Hotels can anticipate the influx of specific traveller types (e.g., more couples in February) and adjust amenities accordingly.
3. **Customer Retention:** The recommendation engine increases engagement by keeping users on the platform longer to discover new attractions.

Future Scope:

- Integrate a **Hybrid Recommendation System** combining content-based (location/type) and collaborative filtering for better accuracy.
- Deploy the model as a real-time API for mobile apps.