

# Assignment: Smart Travel Advisor

---

**Submission:** GitHub Repo + PDF Report

## 1. Project Overview

Build a hybrid Travel Assistant Chatbot that combines:

- RAG (Retrieval-Augmented Generation) for answering travel questions from documents
- Machine Learning for predicting flight/hotel prices
- Intelligent Query Routing to switch between modes

## 2. Detailed Tasks

### Phase 1: Data Preparation

#### A. For RAG System

- Collect 3-5 travel documents (e.g., Lonely Planet guides, airline policy PDFs)
- Preprocess text: Chunk documents (500-800 characters), clean special characters, normalize text
- Generate embeddings using embedding models
- Store in vector DB: FAISS (local) or Chroma

Document	Description	Link (Plain Text)
Travelite India – Brochure	Golden Triangle travel guide (Delhi–Agra–Jaipur), schedule	<a href="https://www.traveliteindia.com/uiimages/downloads/45.pdf">https://www.traveliteindia.com/uiimages/downloads/45.pdf</a>
Air India – Security Regulations	Restrictions on banned items—dangerous goods, electronics	<a href="https://rde.airindia.com/content/dam/air-india/pdfs/security-regulations-dangerous-goods-restricted-items.pdf">https://rde.airindia.com/content/dam/air-india/pdfs/security-regulations-dangerous-goods-restricted-items.pdf</a>
Air India Express – Fees & Charges	Checked baggage allowances, fee schedule for excess weight	<a href="https://airindiaexpress-uat.adobecqms.net/content/dam/air-asia/mandatory-disclosure/AIX-FeesandCharges-12-OCT-23.pdf">https://airindiaexpress-uat.adobecqms.net/content/dam/air-asia/mandatory-disclosure/AIX-FeesandCharges-12-OCT-23.pdf</a>
IndiGo – ZED Travel Policy	Baggage priority, offloading, excess charges for staff travel	<a href="https://cargo.goindigo.in/content/dam/indigov2/slt/ZEDPolicy.pdf">https://cargo.goindigo.in/content/dam/indigov2/slt/ZEDPolicy.pdf</a>
Alliance Air – Baggage Policy	Cabin and check-in baggage limits and dimensions	<a href="https://plone.allianceair.in/allianceair/en/assets/policy/baggage-policy.pdf">https://plone.allianceair.in/allianceair/en/assets/policy/baggage-policy.pdf</a>

## B. For ML Model

- - Use the following dataset for flight price prediction:

<https://www.kaggle.com/datasets/nikhilmittal/flight-fare-prediction-mh>

- Perform EDA (missing values, feature correlations)
- Data Preprocessing
- Train regression model (Random Forest/XGBoost)
- Save model pipeline with joblib including preprocessing steps

## Phase 2: System Implementation

### A. Query Router

Implement logic to detect intent:

- When to call '**RAG**' and when to call '**ML Model**'

### B. RAG Pipeline

- - Use FAISS or Chroma for document retrieval
- - Use Mistral-7B, Llama 3, or OpenAI models for generation

### C. ML Prediction Pipeline

- - Convert user query like 'Predict flight from NYC to London in July' into model features
- - Run prediction using joblib-loaded model
- - Use SHAP/LIME for explanation

## 3. Deliverables

- - GitHub Repo with:
  - - /data: Sample documents + cleaned dataset
  - - /models: Trained ML model + vector store
  - - /notebooks: EDA + model training code
  - - /app: Minimal web app (Streamlit) or notebook demo
  - - Dockerfile: To containerize the app and run locally or in the cloud
- - README.md: Setup instructions + project summary
- - PDF Report (Max 3 pages): Approach, Metrics ( $R^2$ , MAE), Challenges, Improvements

#### 4. Evaluation Rubric

Category	Weight	Criteria
RAG System	45%	Document processing quality, retrieval relevance, response naturalness
ML Pipeline	40%	Feature engineering, model performance, inference robustness
System Design	15%	Query routing logic, error handling, code organization

#### 5. Example Conversations

**User:** What's the best time to visit Bali? (RAG will be called)

**Bot:** According to Lonely Planet (2024), April-October offers dry weather... [source chunk]

**User:** Predict flight price from Paris to Tokyo on December 20 (ML model will be called)

**Bot:** Predicted: \$1,240 (range: \$1,100-\$1,400). Top factors: seasonal demand (+\$300), fuel costs (+\$120)

#### 6. Important Considerations:

- **Time Limit:** Submit your completed assignment, with all code contained within a GitHub repository, latest by **8 PM on July 10, 2025 (Thursday)**. Please include the **link to your GitHub repository** in your submission email.
- **Support:** Please feel free to contact us on email in case of any questions.

**Good luck!**