Implementation Choices & Challenges Report

Project Title: LLM-Powered Booking Analytics & QA System

Implementation Details:

Data Processing:

1. **Data Cleaning:**
   - Handled missing values in columns (children, country, agent, company).
   - Converted reservation_status_date to datetime format.
2. **Feature Engineering:**
   - Calculated total_revenue feature by multiplying adr (average daily rate) with total stay duration from the dataset.
3. **Exploratory Data Analysis (EDA):**
   - Visualized revenue trends over time.
   - Analyzed cancellation rates.
   - Examined the geographical distribution of bookings.
   - Created histograms for lead time distribution.

RAG-based Q&A System:

1. **Sentence Embeddings:**
   - Used all-MiniLM-L6-v2 from sentence-transformers to generate embeddings for text descriptions of bookings.
   - With the help of FAISS Vector Database, Indexed embeddings using faiss.IndexFlatL2.
2. **Retrieval Mechanism:**
   - FAISS searches for the most relevant text snippets based on user queries.
3. **Text Generation:**
   - Used mistralai/Mistral-Small-24B-Instruct-2501 for response generation.
   - Passed retrieved context as input to generate answers.

API Development with FastAPI

1. **Endpoints:**
   - /analytics: Returns cancellation rate and top booking countries.
   - /ask: Processes queries using the RAG model.
2. **Deployment:**
   - Used uvicorn for running the FastAPI server.

Challenges:

1. Model Latency:

- **Issue**: Generative models can have high inference times..

2. FAISS Indexing for Large Datasets:

- **Issue**: Indexing large datasets increases memory usage.

3. Handling Missing Values:

- **Issue**: Missing values in children, country, and company columns.

4. Hugging Face Model Access:

- **Issue**: Requires API token for certain models.

Conclusion:

This project LLM-Powered Booking Analytics & QA System successfully integrates the data analysis, information retrieval, and generative AI to provide insights and an interactive Q&A system for hotel bookings. Future improvements could involve fine-tuning the transformer model on domain-specific datasets for better accuracy.