

TravelIQ- Implementation Report

A Short Report on Implementation Choices & Challenges

Author: Deep Nagpal

Date: 21-03-2025

1. Introduction

- TravelIQ is an AI-powered hotel booking analytics and Q&A system designed to extract insights from hotel booking data.
- It integrates FAISS for vector search and Mistral-7B for AI-generated answers.
- The system is built using FastAPI and provides a RESTful API for querying analytics and booking-related information.

2. Implementation Choices

2.1 Data Processing & Preprocessing

Dataset Used: [hotel_bookings.csv](#)

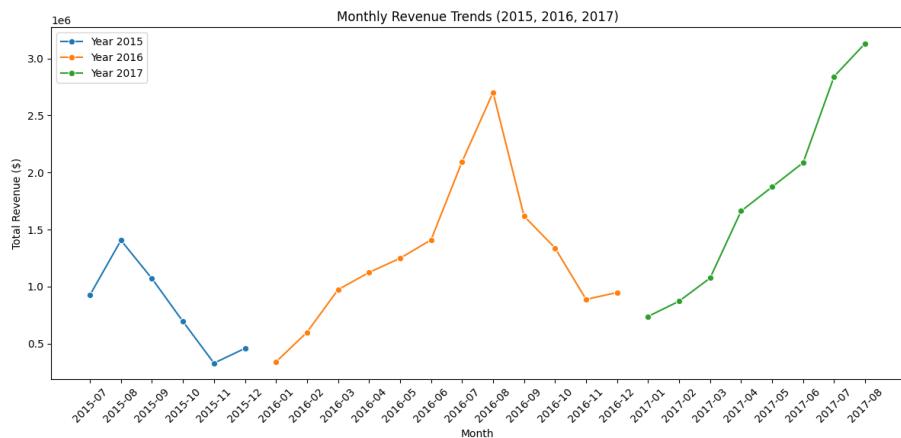
Preprocessing Steps:

- 1 Cleaned missing values, duplicate values and invalid values
- 2 Encoded categorical data (stored as text for FAISS compatibility)
- 3 Created new features (total_revenue, total_nights)
- 4 Converted arrival_date into a structured format
- 5 Stored precomputed analytics in analytics_result.json for fast retrieval

2.2 Analytics & Reporting

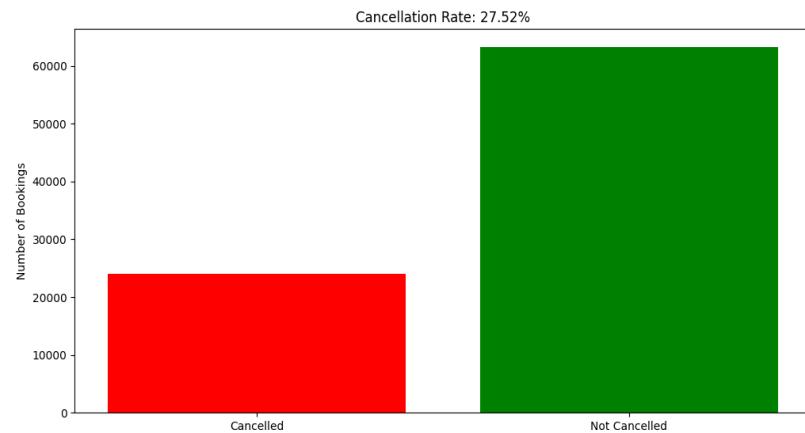
Key Business Insights:

1. Monthly Revenue trends over time



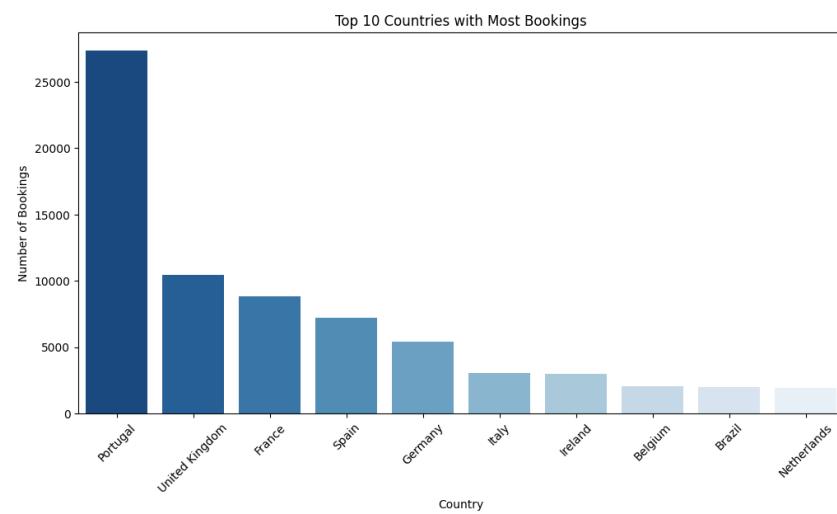
- Revenue trends show clear seasonal fluctuations, with higher earnings during peak travel periods.
- Implementing strategic pricing during low-demand seasons, such as discounts or bundled offers, may help maintain a steady revenue stream throughout the year.

2. Cancellation rate analysis



- The cancellation rate stands at **27.52%**, indicating that over a quarter of bookings are cancelled.
- This suggests potential revenue loss and instability.
- Implementing flexible booking policies, such as refundable rates or discounts for non-refundable bookings, may help reduce cancellations and improve revenue consistency.

3. Geographical distribution

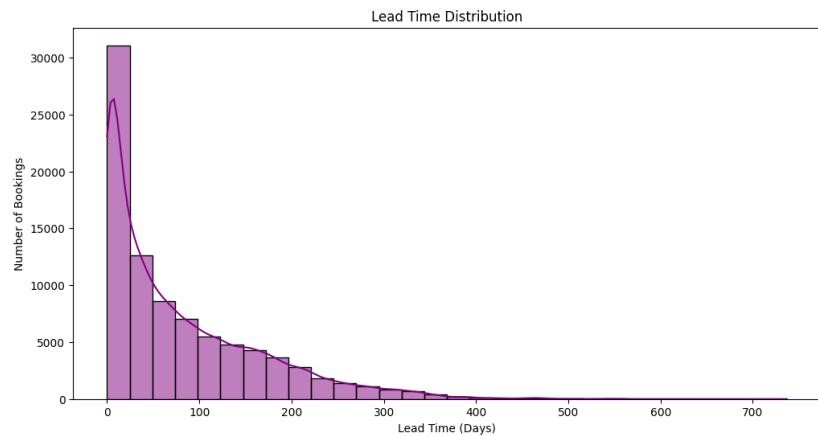


Top 10 Countries with Most Bookings

- Bookings are concentrated in a few top countries, with certain regions dominating the customer base.
- Expanding marketing efforts to underrepresented countries could help increase overall bookings.

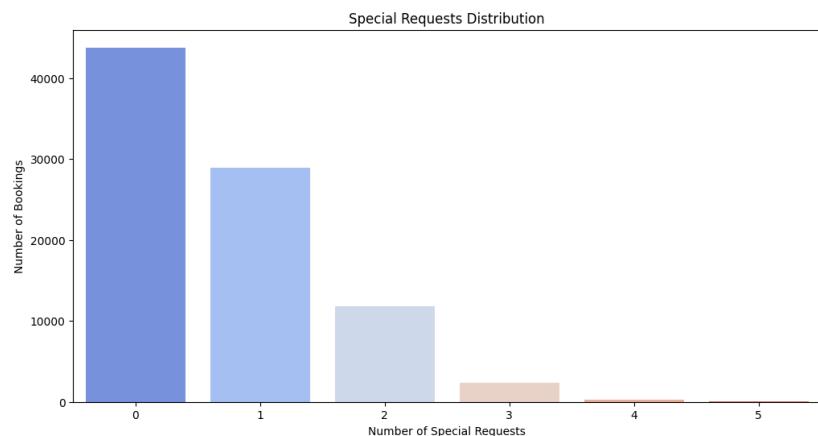
- Tailoring offers to regional preferences may also enhance appeal in untapped markets.

4. Lead Time Analysis



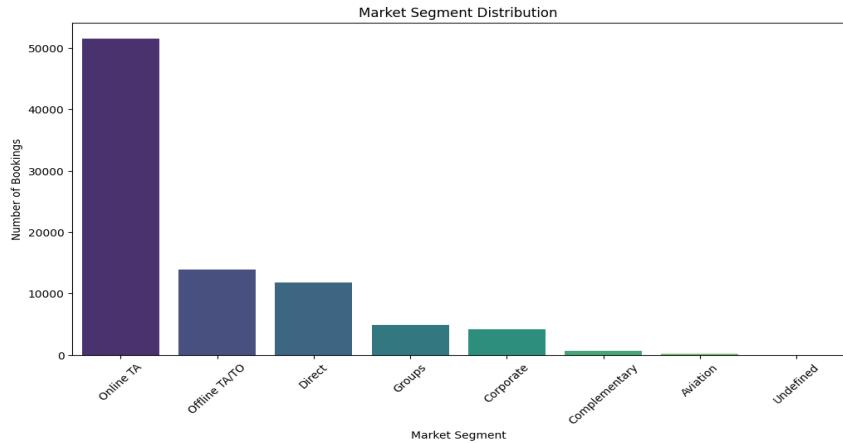
- Lead times primarily range between **0-300 days**, peaking around **100-150 days**, indicating that most bookings are made a few months in advance.
- Offering early bird discounts or incentives for longer lead times could encourage customers to book earlier and improve occupancy planning.

5. Customer Type Distribution



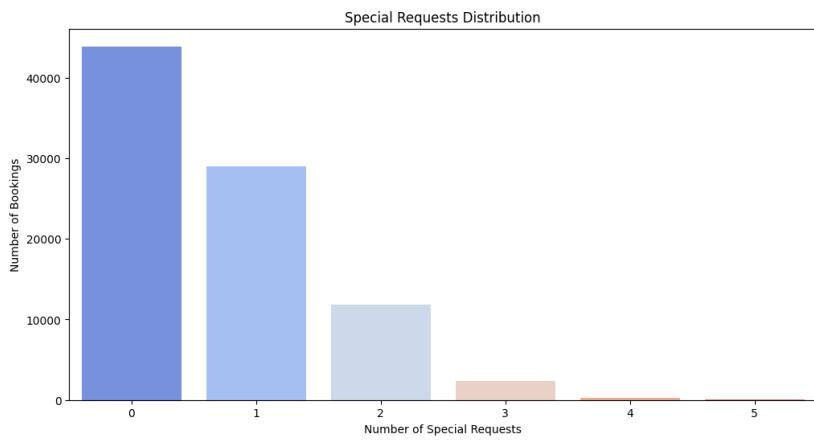
- The majority of bookings come from **Transient** customers, followed by **Transient-Party** and **Contract** types.
- This indicates that individual travellers form the largest customer base.
- Focusing on personalized promotions, loyalty programs, or exclusive offers may improve retention and encourage repeat bookings.

6. Market segment analysis



- Market segments show an uneven distribution, with one dominant category contributing the majority of bookings.
- Diversifying offerings and marketing strategies to attract other segments could create a more balanced customer base and mitigate risks tied to dependency on a single segment.

7. Special Request Analysis



- Most bookings have **0-1 special requests**, indicating limited demand for personalized services.
- This suggests a potential opportunity to introduce additional amenities or customizable packages to enhance customer satisfaction and differentiate from competitors.

2.3 Retrieval-Augmented Generation (RAG) with FAISS

✓ Why FAISS?

- Enables fast and efficient similarity search
- Stores high-dimensional embeddings for rapid query retrieval

📌 Process:

- Extracted key features for embedding
- Used sentence-transformers to generate embeddings
- Stored embeddings in FAISS and implemented top-k retrieval

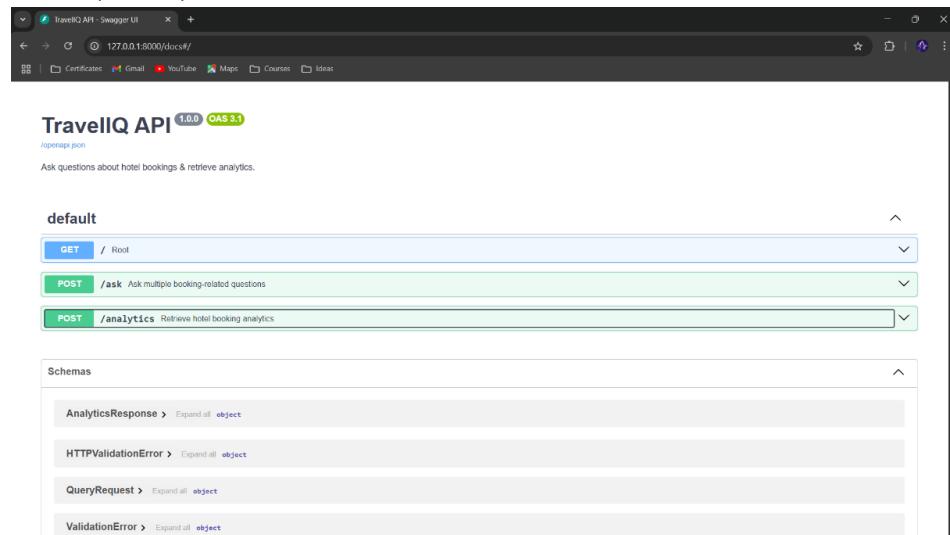
2.4 FastAPI Development

✓ Developed two main API endpoints:

- POST /analytics → Returns precomputed analytics
- POST /ask → Answers booking queries using FAISS + Mistral AI

📌 Additional Features:

- Integrated Hugging Face API for Mistral-7B-Instruct
- Exposed interactive API documentation via Swagger UI (/docs)



3. Challenges & Solutions

Challenges	Solutions
<ul style="list-style-type: none">Large FAISS Index (~127MB) exceeds GitHub's file limit	<ul style="list-style-type: none">Stored faiss_index.bin in Google Drive and provided a download link in README.md
<ul style="list-style-type: none">Hugging Face API returning 404 errors	<ul style="list-style-type: none">Verified API key, switched to HuggingFaceEndpoint, and tested multiple models
<ul style="list-style-type: none">Long API response times for Q&A	<ul style="list-style-type: none">Used precomputed analytics for /analytics to reduce processing time
<ul style="list-style-type: none">Categorical data encoding affecting FAISS retrieval	<ul style="list-style-type: none">Kept categorical columns as text instead of integer encoding

4. Future Enhancements

- Real-time database updates for dynamic insights
- Query history tracking to optimize user experience
- Fine-tuning LLM with hotel-specific data for improved answers
- Expanding FAISS index to support millions of records

5. Conclusion

- TravelIQ successfully integrates analytics, FAISS vector search, and AI-powered Q&A into a single API.
- By combining structured data analysis with intelligent AI-driven responses, TravelIQ provides fast and accurate booking insights.