

# TravellIQ-

# 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](#)

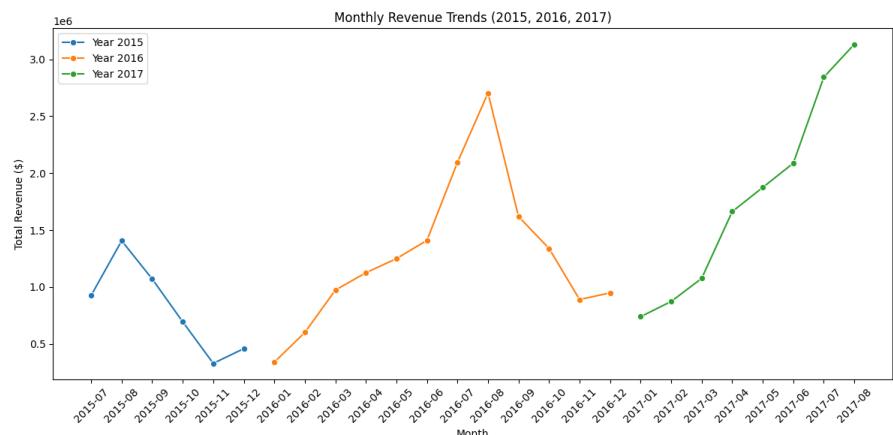
Preprocessing Steps:

- 1 Cleaned missing values
- 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



The line chart illustrates the monthly revenue trends over three years (2015, 2016, and 2017). Each year is represented by a different color:

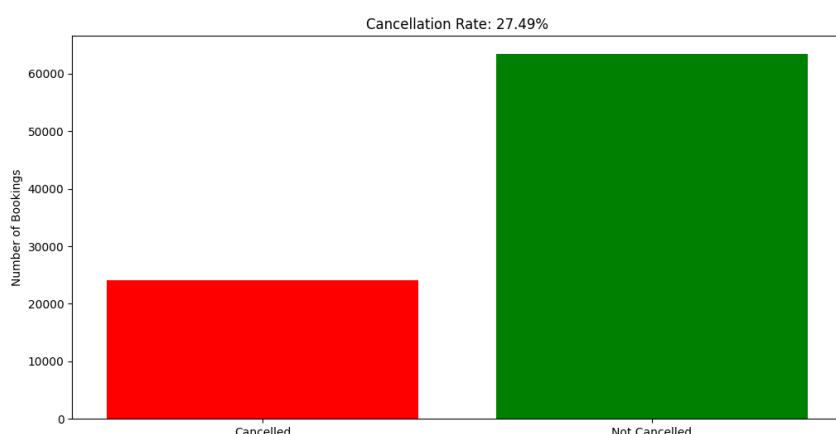
- 2015 (Blue): Revenue fluctuated, peaking around August before declining.
- 2016 (Orange): Revenue showed a steady increase from the beginning of the year, peaking around mid-year (July-August), followed by a decline.
- 2017 (Green): Revenue experienced consistent growth, with a sharp increase after mid-year, reaching the highest value in August 2017.

Key Insights:

- Seasonal Pattern: There is a noticeable peak around mid-year (July-August) across all three years, indicating a seasonal trend in revenue generation.
- Year-over-Year Growth: The revenue trend shows an upward trajectory, with 2017 surpassing previous years in total revenue.
- Fluctuations in 2015 & 2016: Unlike 2017, which shows steady growth, 2015 and 2016 exhibit more fluctuations, especially towards the end of the year.

These insights can help in forecasting future revenue trends and optimizing business strategies for peak seasons.

## 2. Cancellation rate analysis

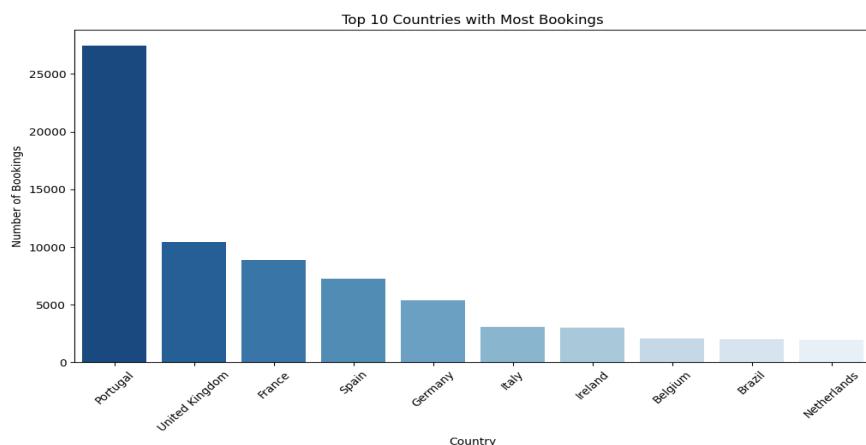


The bar chart displays the number of bookings categorized into Cancelled and Not Cancelled.

#### Key Insights:

- Cancellation Rate: 27.49%, indicating that more than a quarter of total bookings were cancelled.
- Distribution:
  - Cancelled bookings (Red Bar): Around 25,000 bookings were cancelled.
  - Not Cancelled Bookings (Green Bar): More than 60,000 bookings were successfully completed.

### 3. Geographical distribution



#### Top 10 Countries with Most Bookings

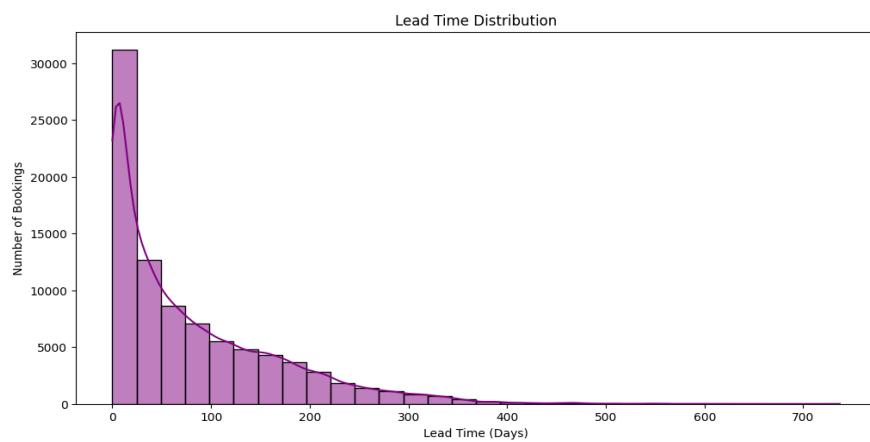
The bar chart illustrates the distribution of hotel bookings across different countries. The data reveals that:

- Portugal dominates the bookings with a significantly higher count than any other country, exceeding 25,000 bookings.
- The United Kingdom, France, Spain, and Germany follow, with a relatively high number of bookings but significantly lower than Portugal.

- Italy, Ireland, Belgium, Brazil, and the Netherlands round out the top 10, with much lower booking numbers.

This indicates that Portugal is a major hub for hotel bookings, possibly due to tourism, business travel, or local factors affecting the hospitality industry. The concentration of bookings in European countries suggests a strong preference for travel within the region.

#### 4. Lead Time Analysis



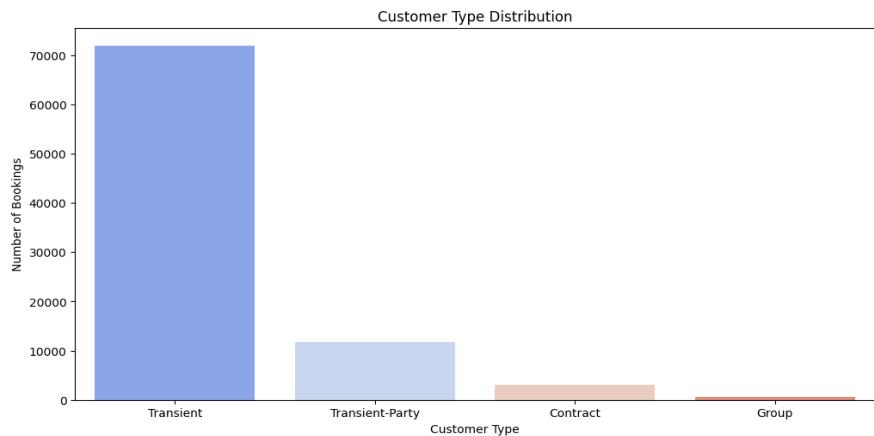
The histogram titled "Lead Time Distribution" provides insights into how far in advance customers book their hotel stays. The key observations from this distribution are:

- A majority of bookings are made with a short lead time, with the highest frequency occurring between 0 to 30 days before the stay.
- As the lead time increases, the number of bookings gradually decreases, forming a right-skewed distribution.
- While some bookings are made several months (even over a year) in advance, such cases are relatively rare.

This trend suggests that most travellers prefer to book their stays closer to their check-in date, which may indicate last-minute travel decisions, spontaneous trips, or price-sensitive behaviour. However, a smaller portion

of travellers still book well in advance, likely due to planned vacations or business trips.

## 5. Customer Type Distribution



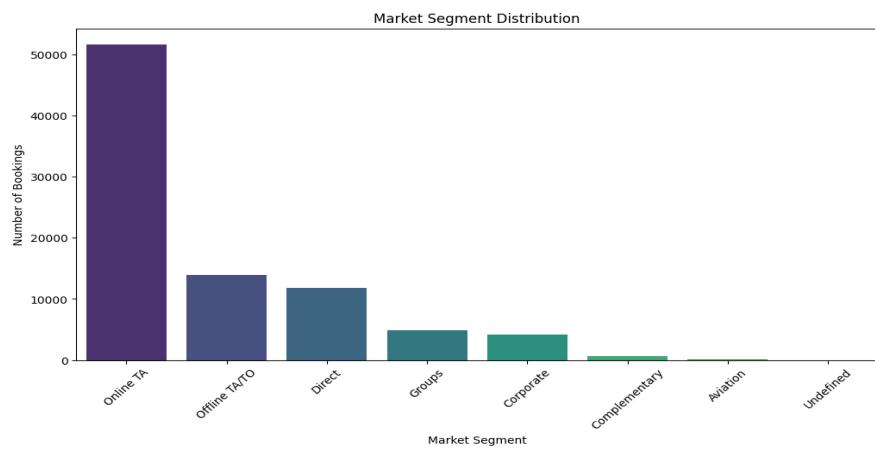
The bar chart titled "Customer Type Distribution" illustrates the number of bookings for different customer types in a hotel dataset. Key observations include:

- Transient customers (individual travellers without contractual obligations) form the largest segment, with over 70,000 bookings. This suggests that most hotel guests are either tourists or business travellers booking short stays.
- Transient-party customers (individuals traveling as part of a small group but booked separately) represent the second-largest category, significantly lower than transient customers.
- Contract customers (those booked through corporate or long-term agreements) form a small portion of total bookings.
- Group bookings (reservations made for large parties, such as conferences or events) are the least common.

## Implications for Hotel Management:

- Since transient customers dominate, hotels should focus on dynamic pricing strategies and last-minute deals to attract individual travellers.
- Promoting group and contract bookings through corporate partnerships or event hosting can help diversify revenue streams.
- Enhancing personalized offers for transient-party customers, such as group discounts or bundled packages, could boost revenue.

## 6. Market segment analysis



The bar chart illustrates the distribution of hotel bookings across various market segments. The x-axis represents different booking sources, while the y-axis shows the number of bookings.

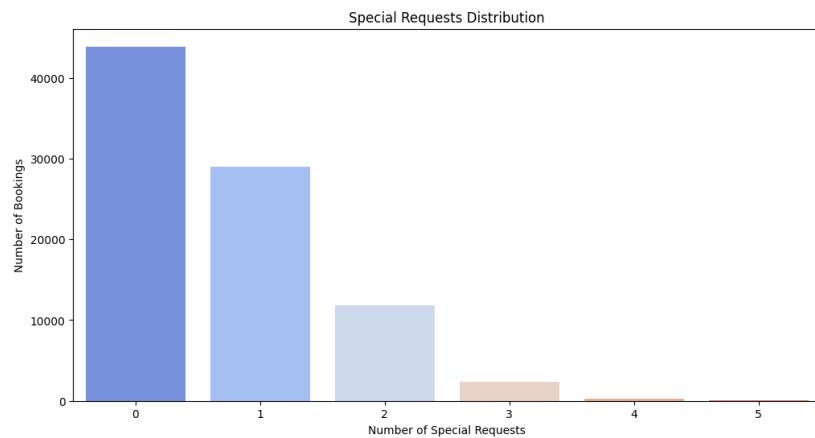
### Key Observations:

- Online Travel Agencies (Online TA) dominate the market, contributing the highest number of bookings (over 50,000).
- Offline Travel Agents/Tour Operators (Offline TA/TO) and Direct Bookings are the next major contributors, with significant but much lower booking counts.
- Groups, Corporate, and Complimentary bookings show relatively fewer reservations.

- Aviation and Undefined categories have minimal representation in the dataset.

This distribution suggests that online platforms are the primary booking channel, making digital marketing and partnerships with OTAs crucial for business growth.

## 7. Special Request Analysis



The bar chart titled "Special Requests Distribution" presents the number of hotel bookings based on the number of special requests made by guests.

Key observations:

- Most guests (over 40,000 bookings) did not make any special requests. This suggests that a large portion of hotel guests are satisfied with standard services.
- Around 30,000 bookings had one special request, and fewer bookings had two or more requests. The number of bookings significantly declines as the number of special requests increases.
- Very few bookings had 4 or 5 special requests, indicating that highly customized requirements are rare.

Implications for Hotel Management:

- Since most guests have no or minimal special requests, focusing on high-quality standard services can enhance overall customer satisfaction.
- Hotels should optimize handling one or two special requests efficiently, as these represent a substantial portion of bookings.
- For the small group of guests with multiple requests, offering a personalized concierge service or premium packages may enhance their experience.

**Tools Used:** Pandas, Matplotlib, Seaborn

**Storage:** Precomputed insights stored in analytics\_result.json

## 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"><li>• Large FAISS Index (~128MB) exceeds GitHub's file limit</li></ul>	<ul style="list-style-type: none"><li>• Stored faiss_index.bin in <b>Google Drive</b> and provided a download link in README.md</li></ul>
<ul style="list-style-type: none"><li>• Hugging Face API returning 404 errors</li></ul>	<ul style="list-style-type: none"><li>• Verified API key, switched to HuggingFaceEndpoint, and tested multiple models</li></ul>
<ul style="list-style-type: none"><li>• Long API response times for Q&amp;A</li></ul>	<ul style="list-style-type: none"><li>• Used precomputed analytics for /analytics to reduce processing time</li></ul>
<ul style="list-style-type: none"><li>• Categorical data encoding affecting FAISS retrieval</li></ul>	<ul style="list-style-type: none"><li>• Kept categorical columns as text instead of integer encoding</li></ul>

## 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.