Infosys Springboard Virtual Internship 6.0 Completion Report

Team Details

Batch Number : 2

Start date : 13-oct-25

Names:

S.no	Name
1	Priya sahu
2	Shruti More
3	Rituraj Suraywanshi
4	Gaurav Raipurkar
5	Ashwin S
6	Manogna Surla

Internship Duration: 8 Weeks

1. Project Title

Revenue Analysis for Hotels

2. Project Objective

The objective of this project is to analyze hotel booking and revenue data using Power BI in order to support data-driven decision-making for hotel management.

- To clean, transform, and structure raw hotel booking data into a reliable analytical model.
- To analyze guest stay behavior, booking patterns, and room utilization across hotel branches.
- To measure hotel performance using key industry metrics such as Occupancy %, ADR, and RevPAR.
- To identify trends in revenue, demand, and cancellations over time.
- To compare booking channels (Direct vs OTA) and evaluate their impact on revenue and commissions.
- To develop interactive dashboards that help revenue managers and general managers monitor performance.
- To support pricing and revenue optimization through forecasting and scenario-based analysis.

Overall, this project aims to convert complex hotel data into clear insights that improve operational efficiency, pricing strategies, and overall profitability.

3. Project description in detail

The AI-Driven Revenue Analysis for Hotels project is designed to analyze hotel booking data and generate meaningful insights that support effective revenue and occupancy management. The project uses Power BI, Power Query, and DAX to transform raw hotel data into interactive dashboards that assist hotel managers in making informed business decisions.

The project begins with collecting and preparing hotel booking data that includes information related to customers, rooms, hotel branches, stay dates, booking duration, pricing, and revenue. Since raw data often contains inconsistencies and missing values, extensive data cleaning and transformation are performed using Power Query. This ensures the accuracy and reliability of all analytical results.

A structured star schema data model is implemented, where the booking table acts as the central fact table and is connected to multiple dimension tables such as Customer, Room, HotelBranch, and Date. This data model enables efficient filtering and aggregation across different dimensions like time, location, and room category.

The project is divided into multiple analytical modules, each focusing on a specific business aspect:

- **Guest Stay Analysis** examines stay duration, stay types (short, medium, long), and room category preferences to understand guest behavior.
- **Occupancy and Revenue Metrics** evaluates hotel performance using key hospitality KPIs such as Occupancy %, ADR (Average Daily Rate), and RevPAR (Revenue per Available Room).
- **Guest Segmentation Analysis** studies customer demographics and booking patterns to identify high-value guest segments.
- **Forecasting and Cancellation Trends** analyzes historical trends to predict future occupancy and understand

cancellation behavior.

- **Revenue Strategy Dashboard** provides interactive tools such as price uplift and occupancy sensitivity sliders to simulate pricing scenarios and optimize revenue.

Interactive dashboards are created to visualize trends at daily, weekly, monthly, and seasonal levels. These dashboards allow users to filter data by branch, room type, season, and booking channel, enabling detailed performance analysis from multiple perspectives.

Overall, this project demonstrates how data analytics and visualization can transform hotel operational data into actionable insights. The solution helps hotel managers optimize pricing strategies, improve occupancy planning, reduce revenue leakage, and enhance overall profitability.

4. Timeline Overview

Week	Activities Planned	Activities Completed
Week 1	Project kickoff, team introduction, defining problem statement, and researching existing customer feedback analysis systems.	Conducted project initiation meeting, finalized objectives and scope, and studied NLP-based sentiment and review analysis approaches.
Week 2	Data collection from review sources (e commerce, app reviews, surveys) and identification of key product feedback datasets.	Collected customer reviews, grouped them by product/category, and validated data formats and sources.
Week 3	Data preprocessing – cleaning text, handling missing fields, and preparing labeled data for sentiment analysis.	Cleaned and normalized review text, performed tokenization and stop-word removal, and created sentiment-labeled datasets for model training.
Week 4	Exploratory Data Analysis (EDA), sentiment distribution analysis, and aspect/category identification from feedback.	Analyzed sentiment trends across products, identified common topics (price, quality, delivery, support), and finalized features for the model.
Week 5	Model development for sentiment analysis and product recommendation.	Trained initial sentiment model and implemented basic recommendation logic using user–product interactions plus sentiment scores.
Week 6	Model testing, evaluation, and fine-tuning of sentiment and recommendation models.	Evaluated models with accuracy/F1 and relevance metrics, tuned hyperparameters, and improved overall system performance. Built an interactive dashboard showing sentiment summaries, top issues, recommended products, and improvement suggestions for product owners. Completed full documentation, prepared presentation slides, and delivered final demo showing end-to-end system workflow from

Week 7	Dashboard and UI creation for user recommendations and product-owner insights (web app / Streamlit / BI tool).	feedback to recommendations and insights.
Week 8	Final documentation, presentation preparation, and submission/demo.	

5a. Key Milestones

Milestone	**Description**	**Date Achieved**
Project Kickoff	Initial project briefing, team formation, objective finalization, and research on customer feedback and sentiment analysis techniques.	6-OCT-25

Milestone	**Description**	**Date Achieved**
Prototype / First Draft Mid-Term Review		

Final Submission Presentation	Completion of data collection, preprocessing, and development of the initial sentiment analysis and recommendation model.	15-OCT-25
	Review of model performance, evaluation results, dashboard progress, and discussions on enhancements for the final system.	30-OCT-25
	Completion of model training, dashboard/interface implementation, documentation, and system deployment for submission.	19-NOV-25
	Final demonstration of the project showing sentiment insights, recommendations, and full dashboard functionality.	01-DEC-25

5b. Project execution details

The AI Skill Resume Matcher & Recommender System was executed over eight weeks using a structured full-stack AI development workflow. The project began with defining objectives, choosing core technologies (Python, Flask, spaCy, scikit-learn, TF-IDF, Cosine Similarity, React.js, Recharts, PostgreSQL, Docker), and planning milestone-based module development with secure JWT authentication. Resumes and job descriptions were uploaded through the backend API and parsed using PyPDF2, python-docx, and custom preprocessing pipelines. NLP-based skill extraction was implemented using spaCy NER with custom skill taxonomies and fuzzy matching for synonym recognition, enabling high-precision identification of both hard and soft skills across multiple domains.

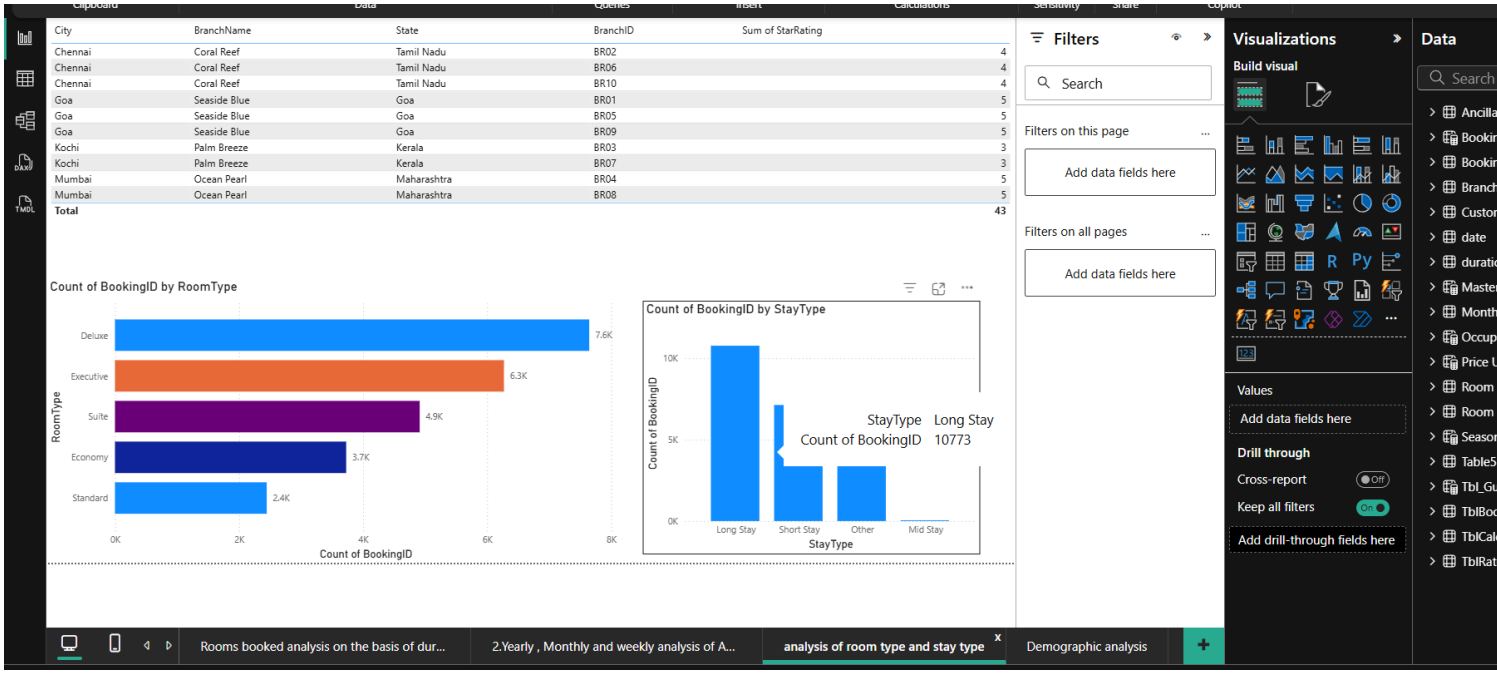
• Extracted skills were compared semantically with job requirements using TF-IDF vectorization and cosine similarity to generate weighted match scores (hard skills 70%, soft skills 30%). The system highlighted matched and missing skills and generated tailored skill-development insights through a recommendation engine that ranked the top skills to learn and mapped career progression paths. All

processed results were stored securely in PostgreSQL and visualized through a responsive React dashboard featuring match breakdowns, skill gap insights, and career roadmap timelines. Final deployment involved containerizing backend and frontend using Docker, configuring CI/CD through GitHub Actions, conducting unit and integration testing (test_skills.py, test_matching.py, test_api.py), and delivering a live demo showcasing the complete user flow from resume upload to personalized recommendations and dashboard visualization.

6. Snapshots / Screenshots

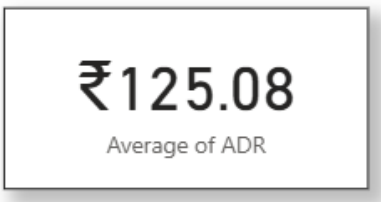
MODULE-1

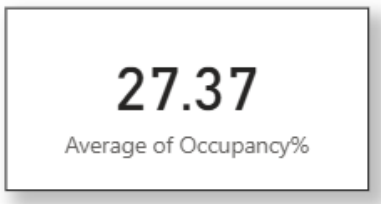
Module 1 focuses on data modeling and ingestion, where raw hotel data is cleaned, structured, and transformed into fact and dimension tables. This module builds a strong data foundation to ensure accurate analysis and seamless reporting in Power BI.

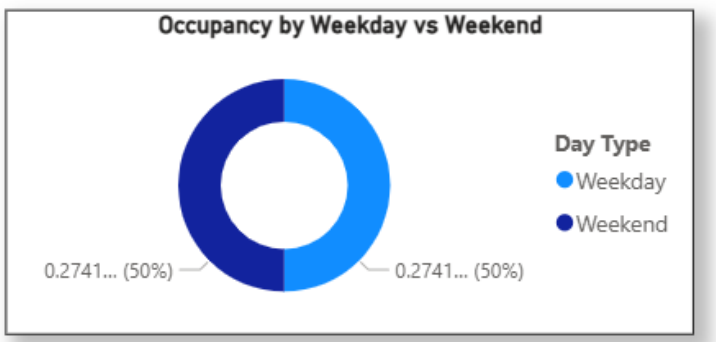


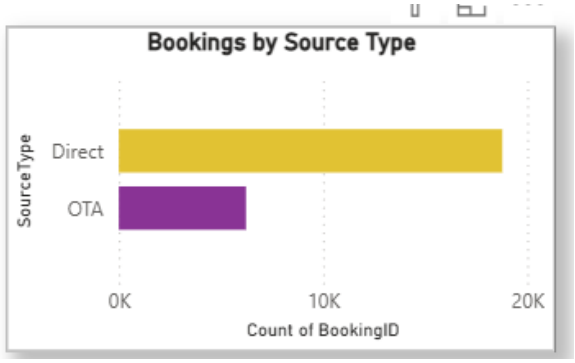
MODULE-2

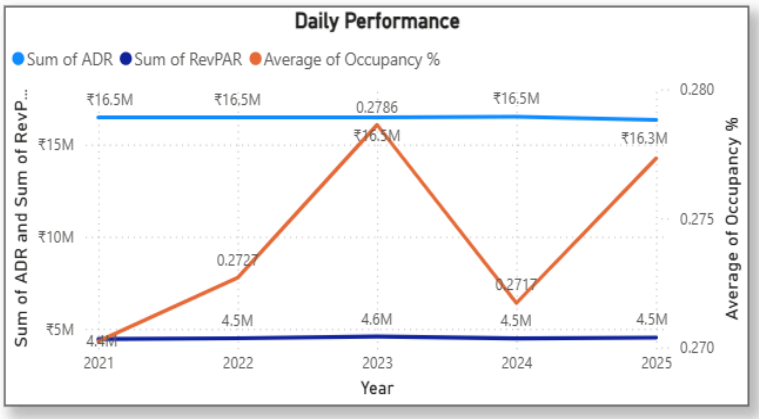
Module 2 focuses on evaluating hotel performance using key occupancy and revenue metrics. This module analyzes how efficiently hotel rooms are utilized, how effective the pricing strategy is, and how well revenue is generated over time. By using industry-standard KPIs such as Occupancy Percentage, Average Daily Rate (ADR), and Revenue per Available Room (RevPAR), along with detailed trend and booking channel analysis, this module supports data-driven revenue and operational decision-making.

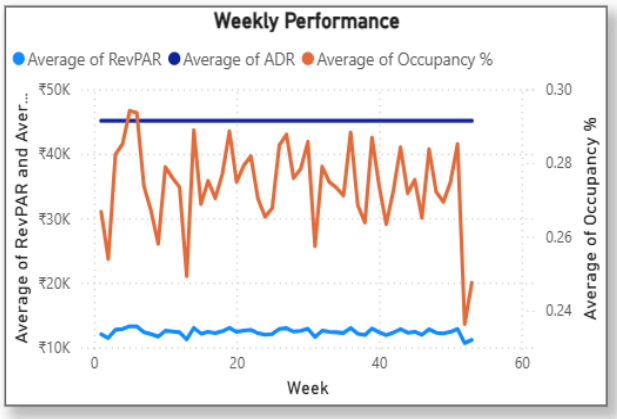


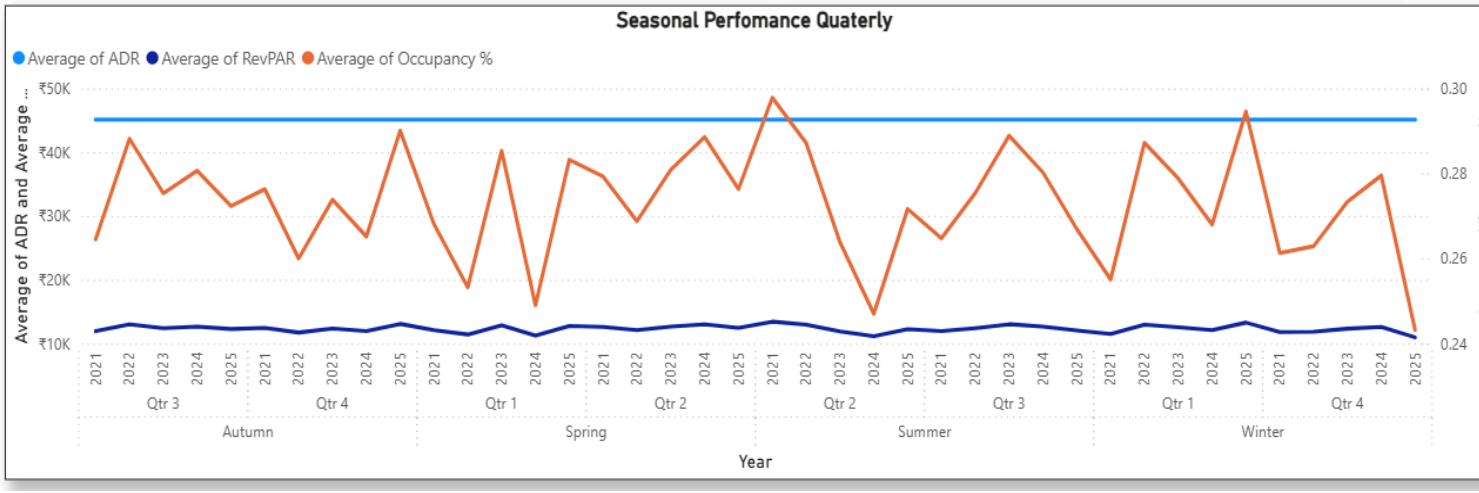






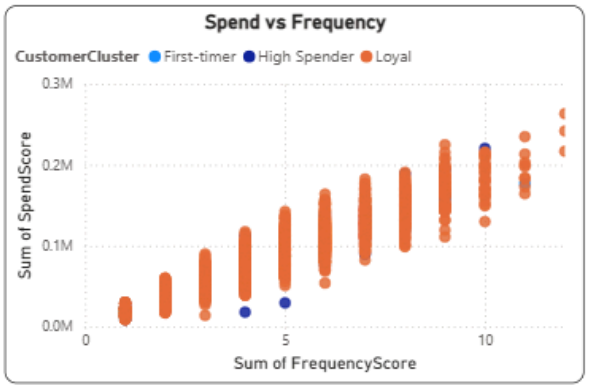


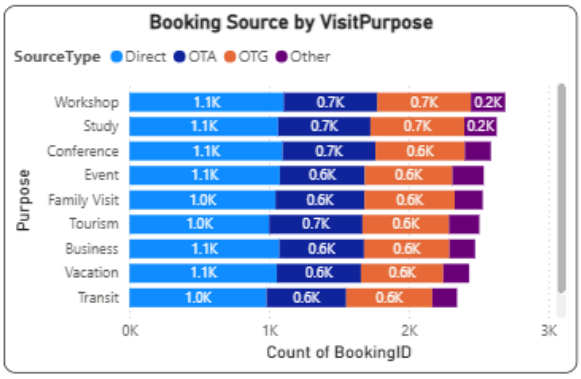




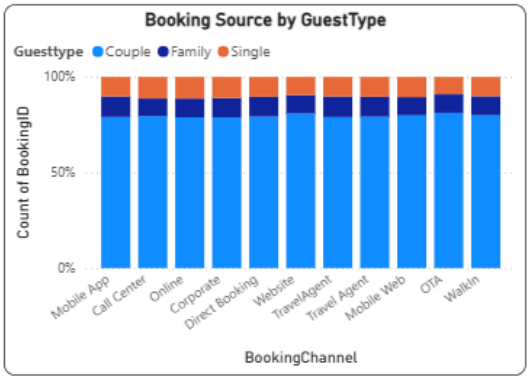
MODULE-3-GUEST ANALYSIS:

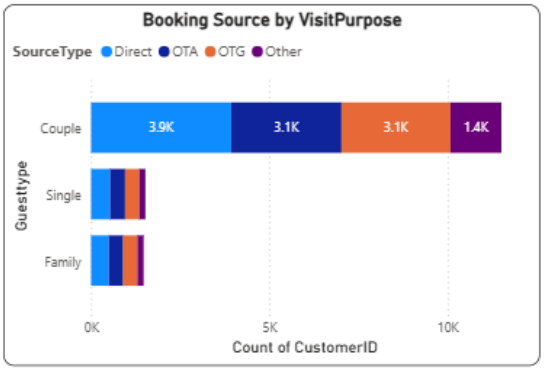
Module 3 analyzes guest behavior and segmentation to identify valuable customer groups. Guests are classified by purpose of visit (Couple, Family, Single), demographics (Child, Teen, Adult, Senior), and booking history (High Spenders, Loyal Guests, First-timers). Using metrics like recency, frequency, and spending, the module highlights top source markets, preferred payment methods, and dominant guest types. These insights help hotels personalize services, improve retention, and optimize revenue strategies.

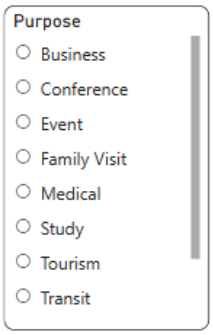

Spend vs Frequency
CustomerCluster First-timer High Spender Loyal
Sum of SpendScore
0.3M
0.2M
0.1M
0.0M
0
5
10
Sum of FrequencyScore

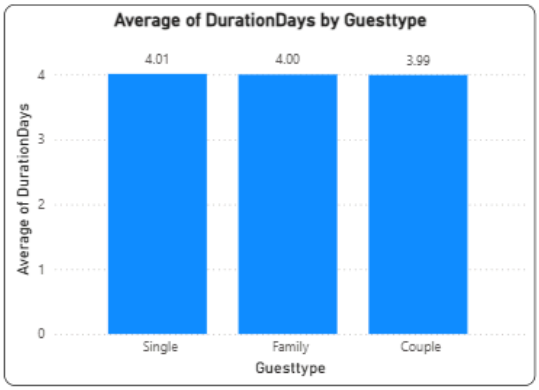

Booking Source by VisitPurpose
SourceType Direct OTA OTG Other
Purpose
Workshop 1.1K 0.7K 0.7K 0.2K
Study 1.1K 0.7K 0.7K 0.2K
Conference 1.1K 0.7K 0.6K
Event 1.1K 0.6K 0.6K
Family Visit 1.0K 0.6K 0.6K
Tourism 1.0K 0.7K 0.6K
Business 1.1K 0.6K 0.6K
Vacation 1.1K 0.6K 0.6K
Transit 1.0K 0.6K 0.6K
0K
1K
2K
3K
Count of BookingID

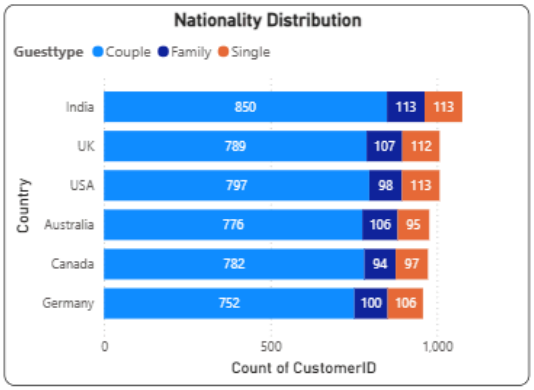

Nationality Distribution
CustomerCluster First-timer High Spender Loyal
EUROPE
ASIA
NORTH AMERICA
AFRICA
AMERICA
AUSTRALIA
SOUTH AMERICA
Pacific Ocean
Atlantic Ocean
Indian Ocean
© 2025 TomTom, © 2025 Microsoft Corporation

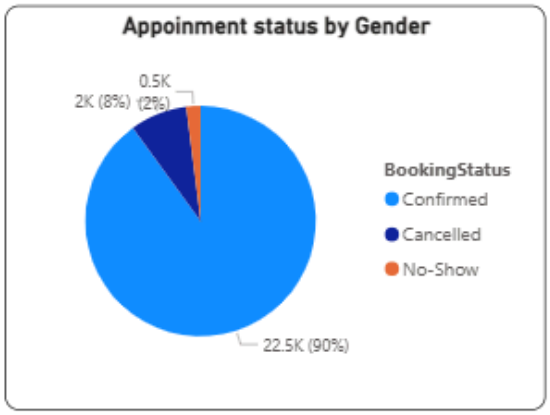

Booking Source by GuestType
Guesttype Couple Family Single
100%
50%
0%
Count of BookingID
Mobile App
Call Center
Online
Corporate
Direct Booking
Website
TravelAgent
Travel Agent
Mobile Web
OTA
Walkin
BookingChannel

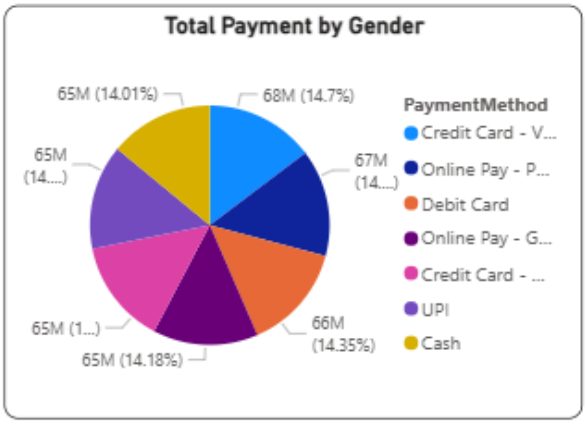

Booking Source by VisitPurpose
SourceType Direct OTA OTG Other
Guesttype
Couple 3.9K 3.1K 3.1K 1.4K
Single
Family
0K
5K
10K
Count of CustomerID

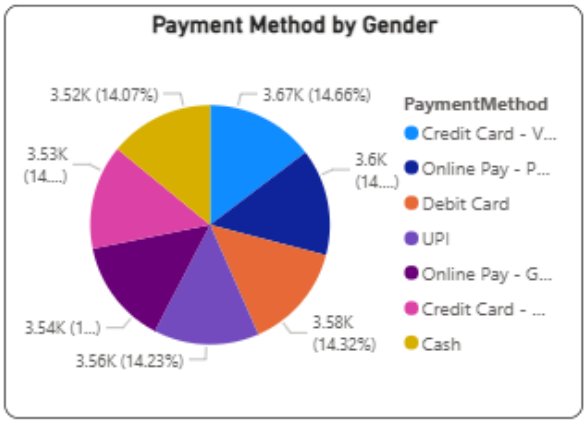

Purpose
Business
Conference
Event
Family Visit
Medical
Study
Tourism
Transit


Average of DurationDays by Guesttype
4.01
4.00
3.99
Average of DurationDays
4
3
2
1
0
Single
Family
Couple
Guesttype


Nationality Distribution
Guesttype Couple Family Single
Country
India 850 113 113
UK 789 107 112
USA 797 98 113
Australia 776 106 95
Canada 782 94 97
Germany 752 100 106
0
500
1,000
Count of CustomerID

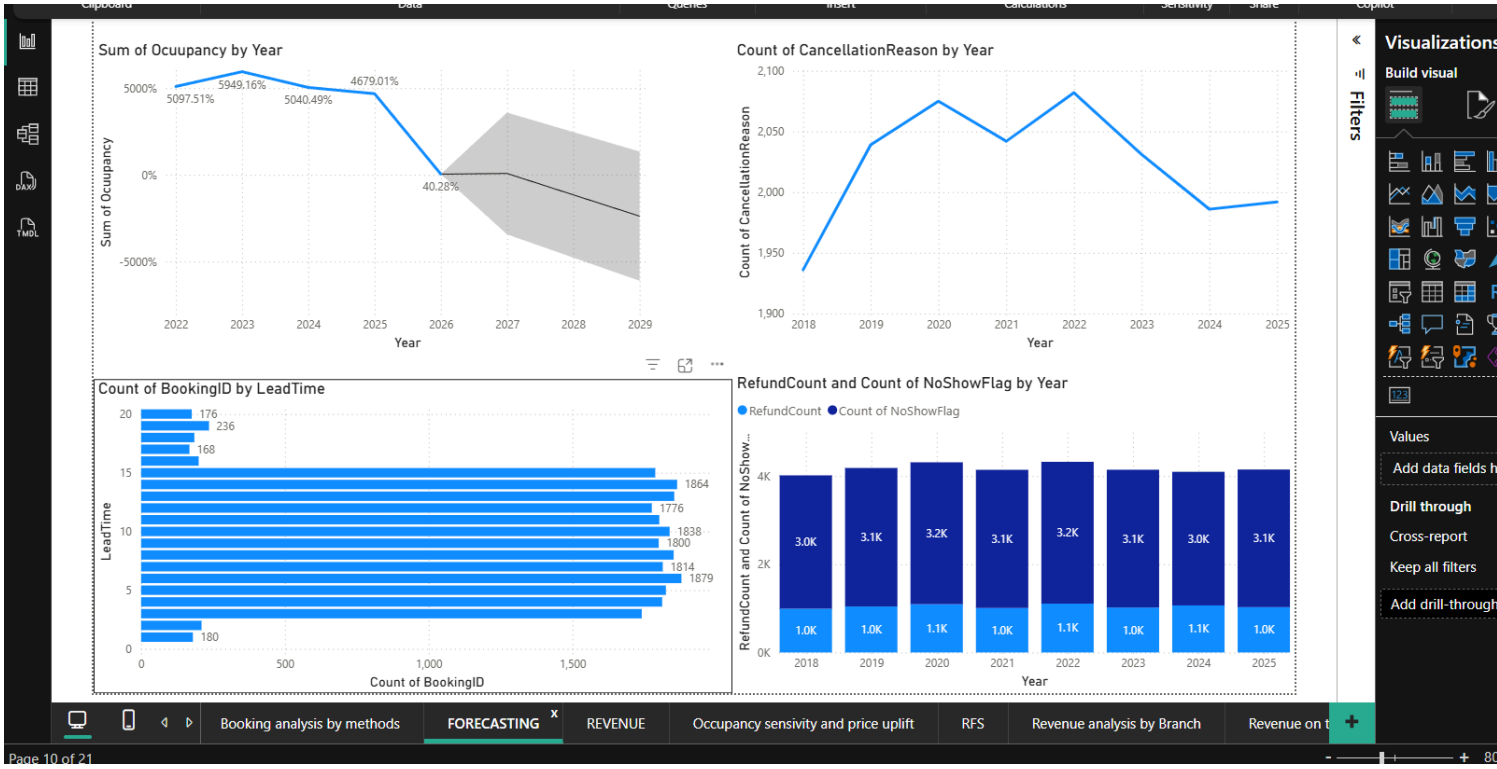


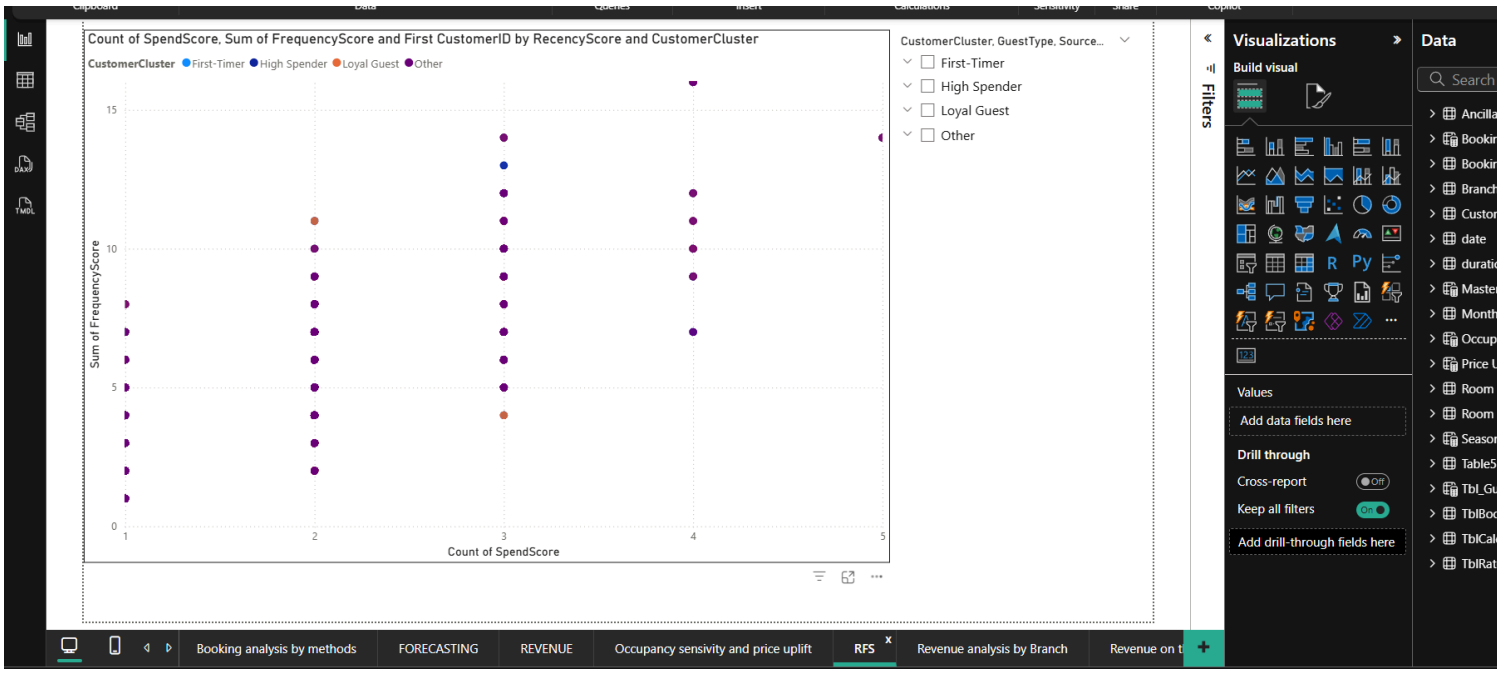




MODULE-4

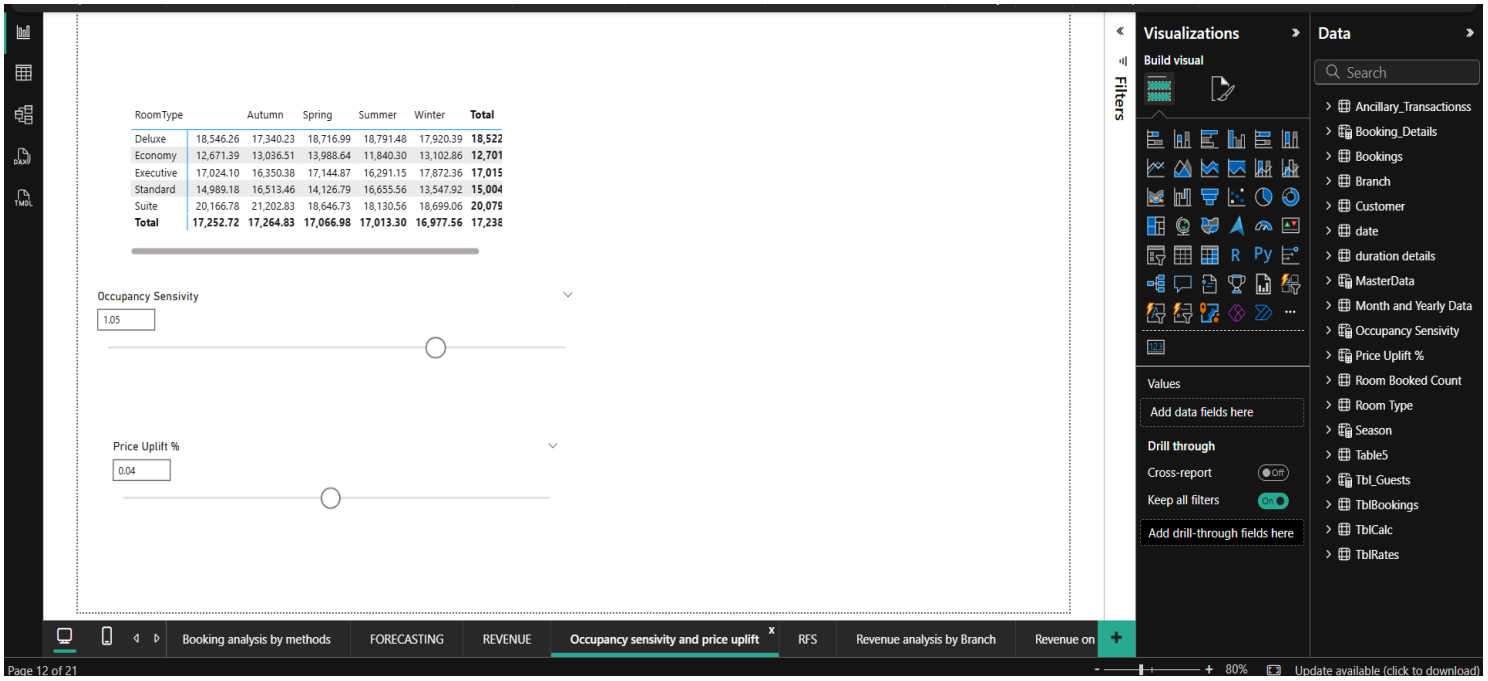
Module 4 focuses on forecasting occupancy and analyzing cancellation behavior to understand future demand trends. It provides insights into lead time, no-shows, refunds, and customer clusters, helping hotels make proactive planning and revenue-protection decisions

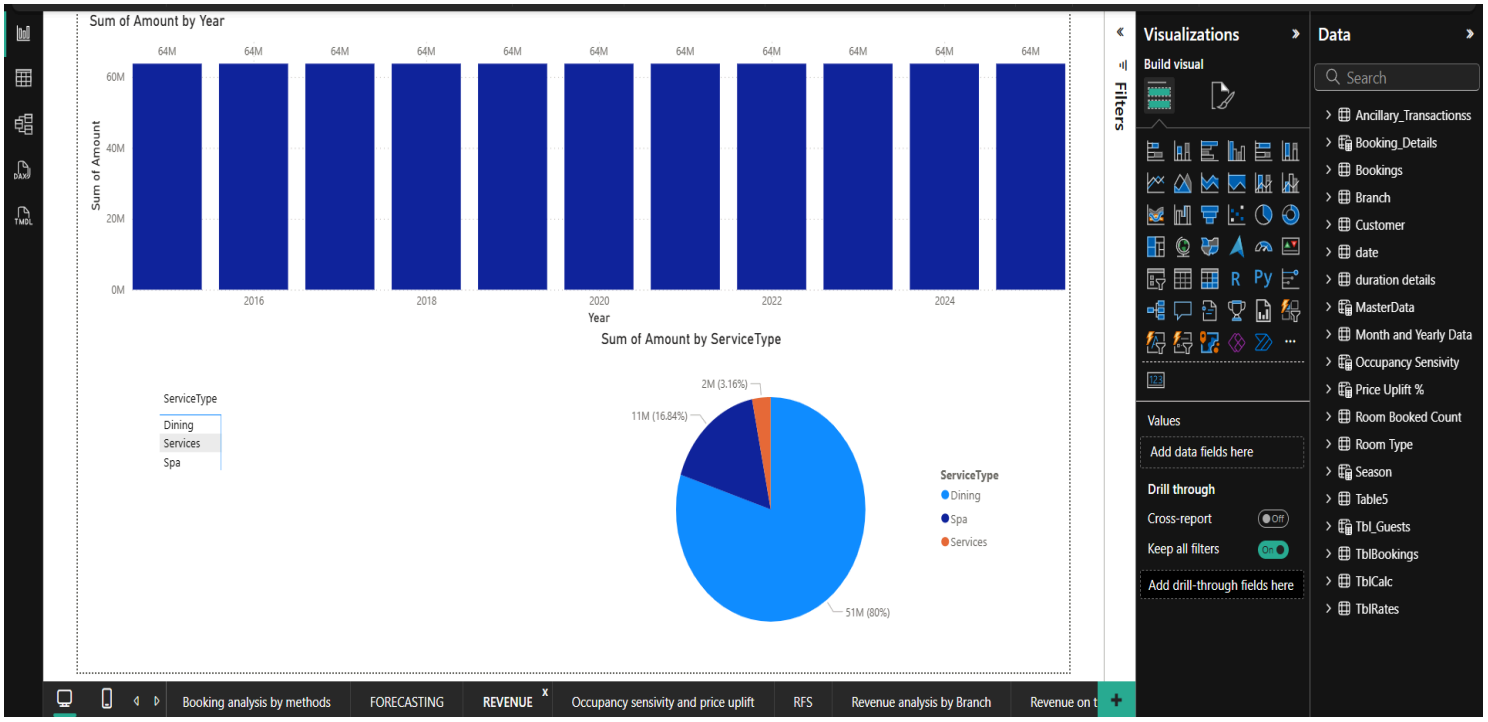




MODULE-5

Module 5 focuses on building an intelligent revenue strategy dashboard with dynamic pricing, occupancy sensitivity, and upsell analysis. It helps revenue and general managers identify pricing opportunities, analyze service-based revenue, and maximize overall hotel revenue through actionable insights.





7. Learnings & Skills Acquired

During the Infosys Springboard internship, I gained strong hands-on experience in **data analytics, business intelligence, and domain-driven problem solving**, specifically within the **hospitality and hotel revenue management domain**. The internship helped bridge the gap between theoretical concepts and real-world business applications.

1. Data Modeling & Warehousing (Star Schema Design)

I learned how to design and implement a **star schema data model** for hotel analytics, which is a foundational

concept in enterprise data warehousing.
The model consisted of a **central fact table (Hotel Bookings)** connected to multiple **dimension tables** such as Date, Room, Customer, and Hotel Branch.

Key learnings included:

- Structuring data for **fast querying and scalability**
- Defining **primary and foreign key relationships**
- Understanding how fact–dimension modeling improves **Power BI performance**
- Mapping business KPIs like **revenue, occupancy, and booking duration** to fact measures

This schema formed the backbone for all dashboards and ensured consistency across modules

2. Hotel Revenue Metrics & Time-Series Analysis

I developed a strong understanding of **hotel performance metrics**, including:

- **Occupancy %**
- **ADR (Average Daily Rate)**
- **RevPAR (Revenue per Available Room)**

By analyzing these metrics **over time** (daily, monthly, and seasonal), I learned how to:

- Identify **seasonal demand patterns**
- Detect pricing inefficiencies (e.g., high occupancy but low ADR)
- Understand the interdependence between Occupancy, ADR, and RevPAR
- Benchmark hotel performance using **competitive analysis concepts** such as RevPAR comparison

This module helped me think beyond raw numbers and focus on **strategic interpretation of trends**

3. Guest Analysis & Customer Segmentation

A major learning area was **guest behavior analysis** and **customer segmentation**, which is critical for personalization and revenue growth in hospitality.

I worked on:

- Classifying guests into **Business, Family, and Solo travelers**
- Analyzing **booking sources** (OTA, Direct, Corporate)
- Studying **stay duration and nationality distribution**
- Segmenting customers into **First-timers, Loyal Guests, and High Spenders**

4. Power BI Dashboard Development & Visualization

The internship significantly improved my **Power BI skills**, including:

- Building **interactive dashboards**
- Designing KPI cards and trend visuals
- Using slicers and filters for multi-dimensional analysis
- Creating matrices and charts aligned with business questions

5. Revenue Strategy & Scenario Simulation (Module 5)

In the final module, I worked on a **Revenue Strategy Dashboard**, which combined analytics with decision-making.

Key learnings included:

- Identifying **upsell potential** across dining, spa, and services
- Understanding **seasonal and room-type pricing strategies**
- Implementing **What-If parameters** such as:
    - Price Uplift %
    - Occupancy Sensitivity
- Performing **scenario analysis** to compare baseline vs adjusted KPIs

6. Business Thinking & Analytical Mindset

Beyond tools and dashboards, the internship helped me develop:

- Business-first analytical thinking
- Ability to translate data insights into **actionable recommendations**
- Understanding of how analytics supports **strategic planning**
- Communication skills to explain insights to non-technical stakeholders

7. Challenges Face

1. Understanding the Hospitality Domain

Initially, understanding hotel-specific concepts such as ADR, RevPAR, occupancy sensitivity, and upsell revenue was challenging. These metrics required domain knowledge beyond basic data analysis.

How I addressed it:

- Studied real-world hotel revenue scenarios

- Mapped metrics to business outcomes
- Used case-based reasoning to understand their impact

---

2. Designing a Scalable Data Model

Creating a star schema that could support multiple analytical modules was complex.
Ensuring correct relationships, avoiding ambiguity, and supporting diverse KPIs required careful planning.

Learning outcome:
This improved my understanding of data architecture and modeling best practices.

---

3. Data Quality & Consistency Issues

Handling missing values, inconsistent booking sources, and irregular date formats was a recurring challenge.

How I handled it:

- Applied data cleaning and standardization techniques
- Validated measures across visuals
- Ensured consistency between fact and dimension tables

---

4. Balancing Simplicity and Insight in Dashboards

A key challenge was deciding what to show vs what to hide in dashboards.
Overloading dashboards with visuals reduces clarity, while oversimplification hides insights.

Learning outcome:
I learned to design dashboards that are insight-driven and user-focused, especially for managerial roles.

---

5. Scenario Modeling & Interpretation

Implementing price uplift and occupancy sensitivity sliders was technically and conceptually challenging.
Interpreting results responsibly without misleading conclusions required careful thinking.

Learning outcome:
This strengthened my ability to simulate business scenarios and evaluate trade-offs.

---

6. Presenting Insights Professionally

Converting technical work into clear explanations for presentations and documentation was initially difficult.

Learning outcome:
I developed better professional communication skills and learned to present analytics from a business perspective.

9.Testimonials from team

During the Hotel Data Analysis internship, our team explored analytical tools and industry-style workflows to study ADR, RevPAR, occupancy, and booking trends. Despite academic schedules, we coordinated through weekly sync-ups, communicated delays proactively, and collaborated to ensure smooth progress. Sharing notes, resolving errors together, and contributing across data preparation and visualization strengthened teamwork and enriched our learning. This collective effort ensured the project's success and built confidence in applying data analytics to real hotel management scenarios.

10. Conclusion

The Hotel Data Analysis project provided a complete, end-to-end learning experience across all major phases of data-driven decision making — from requirement gathering and dataset preparation to statistical modeling, visualization, and dashboard creation. By analyzing key performance indicators such as ADR, RevPAR, occupancy rates, booking sources, cancellation trends, and customer demographics, the project offered valuable insights into hotel operations and revenue management. This hands-on work strengthened practical expertise in data analytics, business intelligence, and visualization tools, while also enhancing strategic thinking around pricing, customer segmentation, and service optimization. Overall, the project played a significant role in building technical confidence and industry readiness, marking an important milestone in the journey toward a career in data science and hospitality analytics.

11. Acknowledgements

I would like to express my sincere gratitude to my mentor for their continuous guidance, encouragement, and technical support throughout this project. Your patience, valuable insights, and motivation were instrumental in deepening my understanding of hotel performance metrics, data visualization, and analytical storytelling. Thank you for believing in my potential and helping me transform this project into a meaningful and successful learning experience.