Proposal Template

1. Project Title

Revenue Optimization Analysis

2. Team Information

- Team Number: 23
- Members: Abhi Sachdeva, Pratham Savjani, Uttam Gowda, Rohith Vinice Richard Arockiaraj
- Course: DSE 501 (Statistics for Data Analytics)
- Semester: Fall 2025

---

3. Dataset Description

The dataset used in this project is the Hotel Booking Demand dataset, which contains detailed booking records for two types of hotels, namely: **a City Hotel and a Resort Hotel**. It contains data collected from **2014 to 2017** and has **119,390 observations and 36 features**, describing both **booking characteristics** and **guest behaviors.**

Data Overview:

- **Hotel type:** city or resort
- **Booking details:** lead time (days between booking and arrival), arrival date, length of stay (weekdays/weekends), number of adults, children, and babies
- **Customer behavior:** previous cancellations, special requests, whether the guest is a repeat, and country
- **Marketing & channel data:** market segment, distribution channel, deposit type
- **Booking outcome:** whether the reservation was canceled, booking status date, and assigned room type.

The dataset has a small amount of missing data, mainly in optional fields such as agent ID, company, and children counts. These will be cleaned or imputed during preprocessing.

Also, it provides almost all the necessary variables to understand and model revenue, pricing behavior, and segment customers.

Potential Audience of the analysis:

Sr No.	**Audience / Stakeholder**	**Why They Find the Analysis Valuable**
1	**Hotel Revenue Managers**	Use data-driven insights to set optimal room prices, forecast demand, and reduce revenue loss from cancellations.
2	**Hotel Owners / General Managers**	Understand overall revenue trends, seasonal profitability, and strategic investment opportunities.
3	**Marketing & Sales Teams (Hotels)**	Identify high-value customer segments and plan targeted promotions or loyalty campaigns.
4	**Online Travel Agencies (e.g., Booking.com, Expedia)**	Predict demand fluctuations and optimize commission structures and listing recommendations.
5	**Travelers & Tourists**	Determine the cheapest times of year to book specific hotels and plan cost-effective vacations.
6	**Corporate & Event Planners**	Forecast accommodation costs and plan conferences or events during low-demand, affordable periods.
7	**Travel-Tech Startups / Hospitality Software Firms**	Build pricing recommendation engines or analytics dashboards (e.g., RoomRev.AI concept).
8	**Tourism Boards & Policy Makers**	Analyze hotel demand seasonality to promote off-season travel and support local economies.
9	**Academic & Research Community**	Use the dataset to teach or study predictive modeling, time series analysis, and business analytics.

Sr No.	**Audience / Stakeholder**	**Why They Find the Analysis Valuable**
10	**Investors & Entrepreneurs**	Identify new opportunities in AI-based hotel management, dynamic pricing, and travel analytics ventures.

4. Objectives & Research Questions

Why the case is important

- Helps identify which factors most affect hotel revenue, such as booking time, season, customer type, and cancellations
- Provides a data-driven approach to improve pricing, marketing, and operational strategies
- Enables more accurate forecasting of demand and revenue
- Supports better decision-making through evidence rather than assumptions
- Reveals customer behavior patterns and booking trends
- Contributes to maximizing occupancy rates and overall profitability
- Shows how statistical analysis and hypothesis testing can solve real-world business challenges
- Builds a bridge between academic research and practical industry applications

Where it is useful

- Hotel revenue management and pricing optimization
- Marketing and customer segmentation strategies
- Demand forecasting and inventory control
- Performance evaluation across different booking channels or customer types
- Planning for seasonal fluctuations and event-based demand
- Improving staff scheduling and operational efficiency

- Guiding investment and expansion decisions for hotel chains
- Applicable to other sectors like airlines, tourism, and event management

---

5. Hypotheses

Fundamental Question: What factors drive revenue, and how can we use them to make better pricing decisions?

- **Hypothesis 1: Seasonal Impact**
  Null Hypothesis - The mean average revenue does not differ significantly across months of the year
- **Hypothesis 2: Family Booking Impact**
  Null Hypothesis - The mean revenue for bookings with children equals the mean revenue for bookings without children
- **Hypothesis 3: Hotel Type Impact**
  Null Hypothesis - The mean revenue for city hotels equals the mean revenue for resorts
- **Hypothesis 4: Customer Type**
  Null Hypothesis - The mean revenue does not differ significantly across customer types
- **Hypothesis 5: Length of Stay Revenue Impact:**
  Null Hypothesis - Total revenue generated per night does not differ significantly across different lengths of stay
- **Hypothesis 6: Year-to-year Revenue Growth**
  Null Hypothesis - The mean revenue generated does not differ significantly across years
- **Hypothesis 7: Market Segment Impact**
  Null Hypothesis - The mean revenue does not differ significantly across market segments

---

6. Methodology & Approach

Phase	Key Tasks	Expected Output
1. Data Acquisition & Exploration	• Load dataset. • Inspect shape, datatypes, summary stats. • Identify categorical/numerical features and target.	• EDA notebook. • Data dictionary with feature info. • Visuals of missing data & distributions.
2. Data Preprocessing & Feature Engineering	• Impute/drop missing values. • Create target: revenue = adr * total_nights. • Engineer features: total_guests, is_family, booking_month, lead_time_group, etc. • Encode categorical variables. • Cap outliers (adr, lead_time).	• Cleaned dataset (hotel_bookings_clean.csv). • Feature pipeline script. • Correlation matrix & updated schema.
3. EDA & Hypothesis Testing	• Analyze revenue, adr, lead_time, cancellations. • Run ANOVA & t-tests (month, hotel, family, etc.). • Visualize p-values & effect sizes. • Logistic regression for cancellation rate.	• EDA report (PDF + Plotly). • Hypothesis summary (p-values, decisions). • Key business insights.
4. Predictive Modeling & Forecasting	• Split data (70/15/15). • Train baseline (Linear, Decision Tree) & advanced (RF, XGBoost, GBM) models. • Evaluate via RMSE, MAE, R^2. • Explain via SHAP. • Forecast monthly revenue (Prophet/SARIMA).	• Trained models (.pkl). • Model report with metrics. • SHAP plots. • 12-month forecast chart.
5. Optimization & Recommendations	• Dynamic pricing simulation. • K-means customer segmentation. • Optional revenue optimization. • Build Streamlit dashboard for visualization. • Compile final insights & presentation.	• Pricing prototype & customer segments. • Interactive Streamlit dashboard. • Final report + slides.