Project Report

Group G-19

Hotel Booking Cancellation & Revenue Optimization Analysis

Sector: ***Tourism & Hospitality***

Team Name: **DVA G-19**

Team Members:

- **Project Lead:** `PRATEEK SHAKYA`
- **Data Lead:** `DAKSH YADAV`
- **Analysis Lead:** `SARTHAK MATHAPATI`
- **Dashboard Lead:** `MOHD. AFFAN ANAS`
- **Strategy Lead:** `YASH RAJ`
- **PPT & Quality Lead:** `SHASHWAT SINHA`
- **Team Assist:** `Vaishnavi Dhanai`

Institute: Newton School of Technology

Faculty Mentor: Archit Raj

Submission Date:18-feb-2026

2.Executive Summary

Problem

Hotels face revenue instability due to high cancellation rates and fluctuating demand patterns. Frequent cancellations directly impact occupancy rates, revenue forecasting, and operational planning. Variations across booking types, customer segments, and seasonal demand further increase uncertainty in financial performance and resource allocation.

Approach:

We analyzed 119,390 booking records using Google Sheets. The analysis included data cleaning, handling missing and inconsistent values, feature engineering, KPI design, and dashboard visualization. Key performance indicators such as cancellation rate, ADR, booking volume, and customer segment performance were evaluated to identify revenue-impacting patterns.

Key Insights

- High lead-time bookings have higher cancellation rates.
- No-deposit bookings show significantly higher cancellation risk.
- Resort hotels generate higher ADR but lower booking volume.
- Certain customer segments drive stable revenue.

Key Recommendations

- Introduce partial deposit for high lead-time bookings.
- Implement dynamic pricing for seasonal peaks.
- Target low-cancellation customer segments for loyalty programs

3. Sector & Business Context

- Tourism contributes significantly to global GDP. Hotels operate under high competition and seasonal demand volatility

Current Challenges:

- High cancellation rates
- Revenue leakage
- Uncertain demand
- Inefficient pricing policies

Why This Problem:

Cancellation directly impacts occupancy and revenue forecasting accuracy.

4. Problem Statement & Objectives

The objective of this project is to analyze booking behavior and identify risk factors influencing cancellations and revenue performance.

Project Scope:

- Analyze historical booking data
- Identify cancellation drivers
- Evaluate revenue patterns
- Design decision-support dashboard

Success Criteria:

- Clear KPIs
- Actionable insights
- Business recommendations
- Interactive dashboard

5. Data Description

Dataset Source

- Dataset: Hotel Booking Demand Dataset
- Source: Kaggle
- Access Link: https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand
- Dataset: Hotel Booking Demand (Jesse Mostipak)
- Data Structure:119,390 rows,32 columns

Booking-level data

Key Columns:

- is_canceled
- lead_time
- hotel
- market_segment
- deposit_type
- adr

Limitations:

- No real-time updates
- Country codes abbreviated
- No detailed cost structure

6.Data Cleaning & Preparation

#Missing Values Handling:

The dataset contained missing values in several columns, particularly:

- `Agent`
- `Company`
- `Children`
- `country`

Actions Taken:

- Missing values in **agent** and **company** were replaced with `0` to indicate no travel agent or corporate booking.
- Missing values in **children** were replaced with `0`, assuming no children if not specified.
- Missing values in **country** were labeled as “Unknown” to preserve booking records while maintaining category consistency.

#Outlier Treatment

The `adr` (Average Daily Rate) column contained extreme values that could distort revenue analysis.

Actions Taken:

- Identified unusually high ADR values using sorting and filtering.
- Reviewed outliers and retained realistic values while ensuring no data-entry errors skewed calculations.
- Extreme ADR values beyond logical hospitality pricing ranges were reviewed carefully before inclusion.

#Transformations

Total Stay Nights=`stays_in_weekend_nights + stays_in_week_nights`

Total Guests= `adults + children + babies`

Total Revenue=`adr × Total Stay Nights`

#Feature Engineering

Feature engineering was conducted to enhance business insights:

- Created **Lead Time Buckets** (0–30, 31–90, 91+ days)

- Created **Cancellation Flag** using is_canceled
- Derived revenue-based metrics for segmentation

#Assumptions

- Missing children values were assumed as zero.
- ADR reflects per-night price and accurately represents booking revenue.
- Historical data patterns are representative of operational trends.

#Tool Compliance

All cleaning, transformations, and derived columns were executed in **Google Sheets**, adhering strictly to capstone requirements.

7. KPI & Metric Framework

To evaluate cancellation risk and revenue performance, a structured KPI framework was developed. Each metric directly aligns with the project objective: **reducing cancellations and maximizing revenue stability**.

#Cancellation Rate (%)

Definition:
Percentage of total bookings that were canceled.

Formula (Google Sheets):
`=AVERAGE(is_canceled)`
(Where 1 = Canceled, 0 = Not Canceled)

Why It Matters:
Measures booking volatility and revenue risk. High cancellation reduces predictability and affects occupancy planning.

Objective Mapping:
Identifies structural booking instability and supports risk-reduction strategy.

#Total Revenue

Definition:
Total income generated from completed bookings.

Formula:
`Total Revenue = ADR × Total Stay Nights`
Where:
Total Stay Nights = stays_in_weekend_nights + stays_in_week_nights

Why It Matters:
Primary indicator of financial performance and pricing effectiveness.

Objective Mapping:
Measures success in revenue maximization.

#Lost Revenue (Due to Cancellations)

Definition:
Revenue forfeited from canceled bookings.

Formula:
`Lost Revenue = ADR × Total Stay Nights (for canceled bookings)`

Why It Matters:
Quantifies direct financial impact of cancellation behavior.

Objective Mapping:
Measures revenue leakage and supports deposit policy decisions.

#Average Daily Rate (ADR)

Definition:
Average revenue earned per room per night.

Formula:
`=AVERAGE(adr)`

Why It Matters:
Indicates pricing strategy efficiency and market positioning.

Objective Mapping:
Supports dynamic pricing and profitability optimization.

#Revenue by Segment

Definition:
Revenue contribution across customer types and market segments.

Method:
Pivot table using `SUM(Total Revenue)` grouped by segment.

Why It Matters:
Identifies high-value, low-risk segments for strategic focus.

Objective Mapping:
Enables targeted marketing and revenue-quality improvement.

8. Exploratory Data Analysis (EDA)

Exploratory Data Analysis was conducted to identify structural patterns affecting cancellation risk and revenue performance across hotel bookings. The analysis was organized across trend, comparison, distribution, and relationship dimensions to derive actionable insights.

Trend Analysis

Monthly booking patterns reveal clear seasonality, with peak demand observed during mid-year periods. Resort hotels demonstrate stronger seasonal fluctuations compared to city hotels, indicating higher sensitivity to tourism cycles. Cancellation rates also vary across months, suggesting that higher booking volume does not always translate into stable revenue. These findings highlight the importance of season-adjusted revenue forecasting and dynamic pricing strategies.

Comparison Analysis

Cancellation behavior differs significantly across booking conditions. Bookings without deposit requirements show substantially higher cancellation rates compared to non-refundable or partially refundable bookings, indicating that financial commitment reduces cancellation risk. Additionally, city hotels exhibit higher cancellation rates than resort hotels, suggesting greater volatility in business travel segments. Revenue comparison across market segments reveals that certain customer categories

contribute higher revenue with relatively lower cancellation risk, emphasizing the need for targeted customer strategies.

Distribution Analysis

Lead time distribution is positively skewed, with a notable proportion of bookings made well in advance. Longer lead times are associated with increased cancellation likelihood, reflecting greater uncertainty over extended booking horizons. The Average Daily Rate (ADR) distribution is moderately concentrated within mid-range pricing bands, with limited extreme outliers, indicating stable pricing behavior with selective premium segments contributing higher revenue.

Correlation & Relationship Insights

A positive association exists between lead time and cancellation probability, confirming that advance bookings carry higher cancellation risk. Deposit type demonstrates a strong inverse relationship with cancellation, where non-refundable bookings significantly reduce revenue volatility. These relationships indicate that cancellation risk is structurally influenced by booking behavior rather than random variation.

10. Dashboard Design

The dashboard was fully developed in **Google Sheets** using pivot tables, calculated fields, derived columns, and interactive slicers. All analysis and transformations were executed within the master sheet to comply with capstone requirements.

Objective

The dashboard serves as a decision-support tool to:

- Monitor booking and revenue performance
- Identify key cancellation risk drivers
- Analyze seasonal and segment-level patterns

- Support strategic revenue optimization

Structure

The dashboard follows a top-down executive structure:

1) KPI Summary Layer

Displays:

- Total Bookings
- Total Cancellations
  Total Revenue
  Lost Revenue
- ADR

2) Trend Analysis

Includes:

- Bookings by Year
- Monthly Booking vs Cancellation Trend

3) Risk Driver Analysis

Includes:

- Cancellation Rate by Hotel Type
- Lead Time Distribution
- Cancellation vs Deposit Policy
  Purpose: Identify controllable cancellation drivers.

4) Revenue Segmentation

Includes:

- Revenue by Customer Segment
  ADR by Customer Type

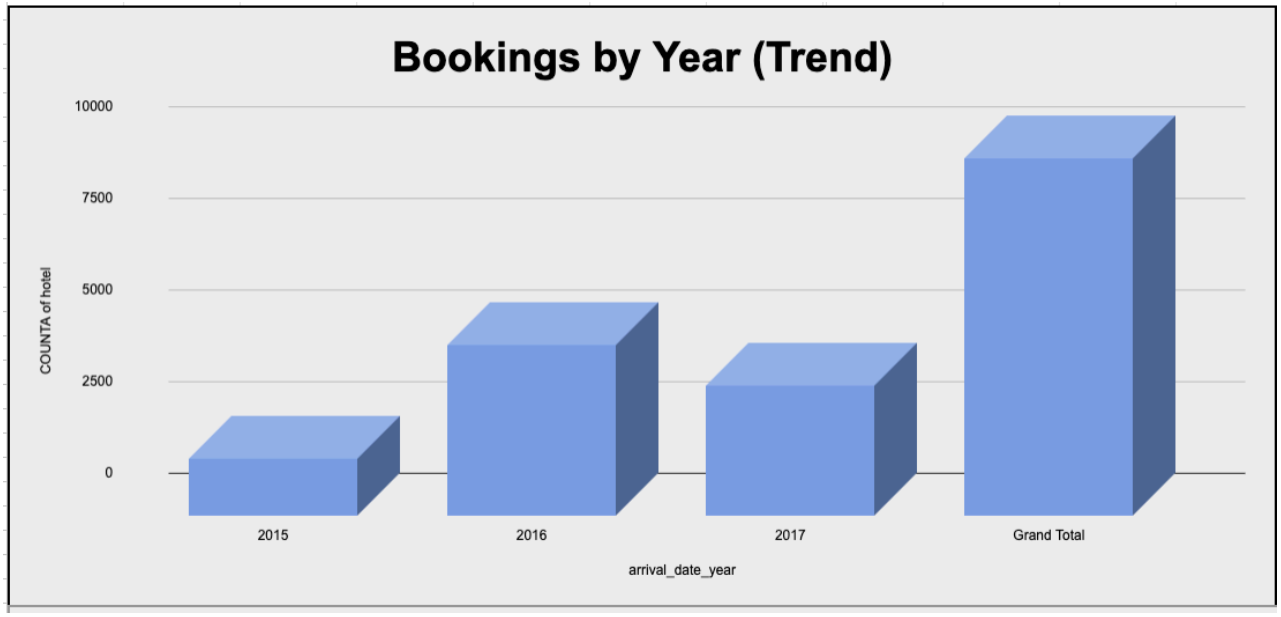
Lost Revenue by Market Segment

Purpose: Identify high-value and high-risk segments.

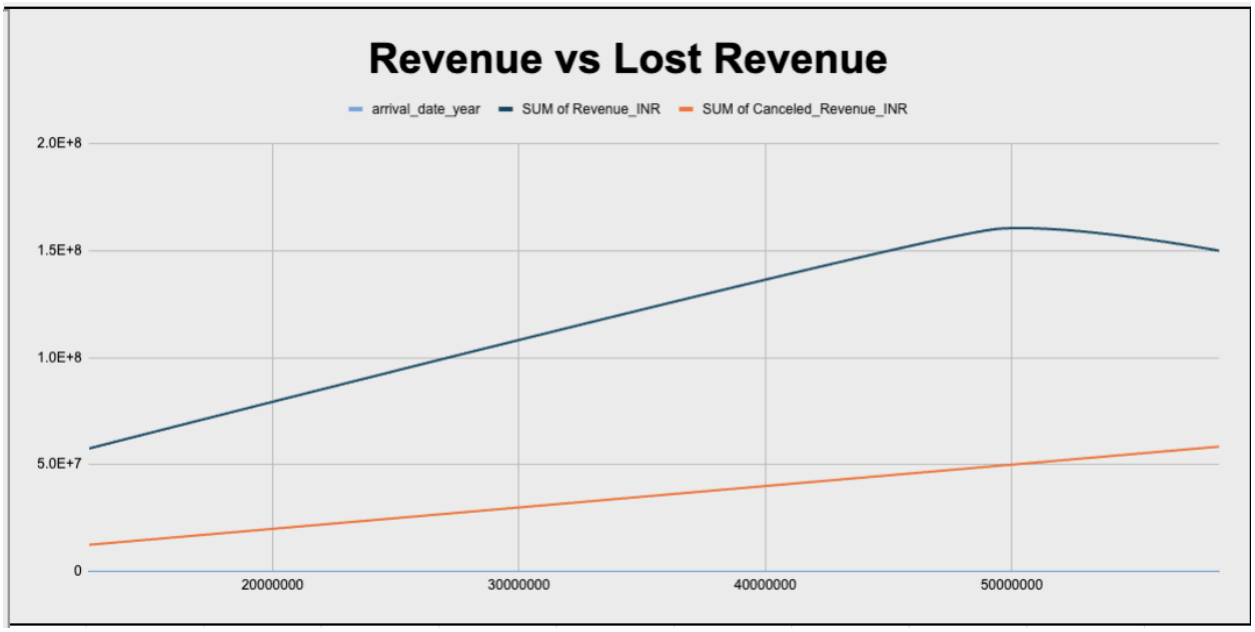
Filters & Interactivity

Interactive slicers allow filtering by:

- Arrival Year
- Customer Type
- Risk Drivers



KPIs: Total Bookings and Year-on-Year Growth.
Use: Analyzes annual demand trends, growth patterns, and supports forecasting and revenue planning decisions.



KPIs: Total Revenue and Lost Revenue (Canceled Revenue).
Use: Evaluates financial performance, revenue leakage, and impact of cancellations on profitability trends.

11. Insights Summary

The analysis reveals the following key insights:

1. Cancellation rates are materially high, posing structural revenue risk.
2. No-deposit bookings exhibit significantly higher cancellation probability.
3. Longer lead times are strongly associated with increased cancellation risk.
4. City hotels show greater cancellation volatility than resort hotels.
5. Certain market segments deliver stable, high-quality revenue with lower cancellation exposure.
6. Seasonal peaks increase booking volume but also amplify cancellation uncertainty.
7. ADR remains concentrated in mid-range pricing bands, with premium segments driving disproportionate revenue.
8. Repeat guests demonstrate lower cancellation risk, indicating loyalty-driven stability.

9. Higher special-request bookings reflect stronger commitment and reduced cancellation probability.

10. Deposit policy and lead-time management emerge as the most controllable drivers of revenue protection.

---

12. Recommendations

1. Tiered Deposit Policy

Insight: No-deposit bookings drive higher cancellations.
Action: Introduce partial deposits for bookings with long lead times.
Impact: Improves revenue predictability and reduces cancellation volatility.
Feasibility: High.

2. Risk-Based Segmentation

Insight: Long lead-time bookings increase risk.
Action: Implement booking risk scoring using lead time and segment variables.
Impact: Enhances forecasting accuracy.
Feasibility: Medium.

3. Focus on Stable Segments

Insight: Certain segments provide high revenue stability.
Action: Reallocate marketing efforts toward low-risk, high-value segments.
Impact: Improves revenue quality.
Feasibility: High.

4. Dynamic Seasonal Pricing

Insight: Clear seasonal demand fluctuations.
Action: Apply demand-based pricing adjustments.
Impact: Maximizes peak-period revenue.
Feasibility: High.

5. Strengthen Loyalty Programs

Insight: Repeat guests show lower cancellation risk.
Action: Enhance retention incentives.
Impact: Improves booking stability.
Feasibility: Medium.

---

13. Impact Estimation

Logical projections suggest:

- A 5% reduction in cancellation could materially improve effective occupancy and revenue stability.
- Deposit enforcement can reduce revenue leakage from high-risk bookings.
- Segment-focused marketing increases return on marketing spend.
- Dynamic pricing improves ADR during peak demand.
- Improved booking predictability reduces administrative and operational inefficiencies.

Collectively, implementation can:

Reduce financial risk
Improve revenue stability
Enhance planning efficiency
Strengthen service consistency

---

14. Limitations

- Analysis is based on historical data only.
- Operational cost data was unavailable, limiting profitability analysis.
- External economic or policy shocks are not captured.
- Behavioral drivers are inferred from booking data without demographic enrichment.
- Observed correlations do not imply direct causation.

---

15. Future Scope

- Incorporate real-time booking data.
- Develop predictive cancellation models.
- Integrate customer sentiment and review analytics.
- Include competitive pricing benchmarks.
- Calculate RevPAR with cost integration.
- Apply machine learning for risk classification.

---

16. Conclusion

This project transformed raw booking data into structured, decision-oriented intelligence. The analysis confirms that cancellation risk and revenue performance are primarily influenced by deposit policy, lead time, seasonality, and customer segmentation. Strategic interventions in these areas can significantly enhance revenue stability and operational efficiency. The developed dashboard provides a practical decision-support framework for proactive revenue management.

---

18. Contribution Matrix (Mandatory)

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Prateek Shakya	✔	✔	-	✔	✔	✔	Project Lead
Yash Raj	-	✔	✔	✔	-	-	Strategy Lead
Sarthak Mathpati	✔	✔	-	✔	-	-	Analysis Lead
Affan anas	✔	-	✔	✔	-	-	Dashboard Lead
Shashwat Sinha	✔	-	✔	-	✔	✔	Ppt & quality Lead
Daksh yadav	✔	✔	✔	-	✔	-	Data Lead
Vaishanvi	-	-	-	-	✔	✔	Team Assist

Declaration: We confirm that the above contribution details are accurate and verifiable through version history and submitted artifacts.

Team Signature Block: _All is well_____