Capstone Project Report

1.Cover Page

Project Title:

 Hotel Booking Performance & Cancellation Intelligence Dashboard

Sector:
 Hospitality & Tourism Analytics

Team Members:

- Kasula Lalithendra - 2401010216 (Analysis & Strategy Lead)
- Abhiman SIngh - 2401010016 (Data Lead)
- Vridhi Chaudhary - 2401010336(Project Lead)
- Ritik Raj - 2401010387 (Strategy Lead)
- Ananth Singh - 2401010067 (PPT & Quality Lead)
- Rudraksh Sharma - 2401010395 (Dashboard Lead)

Institute : Newton school of Technology

Course:
 Data Visualization & Analytics Capstone   Submission Date:18/02/2026

2.Executive Summary

Business Problem

The hospitality industry faces significant revenue volatility due to high booking cancellation rates, seasonality, and unpredictable customer behavior. The dataset reveals an overall cancellation rate of approximately **32.6%**, indicating that nearly one in three bookings does not convert into actual revenue realization.

This creates:

- Revenue instability

- Inefficient inventory planning

- Operational uncertainty

- Ineffective deposit strategy

- Marketing misalignment

Project Objective

To design an **interactive decision-support dashboard** that analyzes:

- Hotel Type performance (City vs Resort)

- Market Segment cancellation risk

- Distribution channel profitability

- Lead time impact on cancellation

- Deposit policy effectiveness

- Repeat guest behavior

- Seasonal revenue trends

Approach

- Data cleaning & transformation in Google Sheets

- Pivot-based KPI modeling

- Feature engineering (Lead Time Groups, Revenue calculation)

- Interactive slicer-controlled dashboard

- Comparative and risk analysis

Key Insights

- Group bookings show the highest cancellation rate **(~54%)**

- 180+ day lead time bookings are the most unstable **(~53% cancellation)**

- Non-Refund deposit type shows **unusually high cancellation (~98%)** indicating possible policy/data issue

- Revenue peaks mid-year **(Month 7–8)**, showing strong seasonality

- Corporate and Direct bookings are more **stable** than **OTA channels**

Key Recommendations

- Implement **stricter policies** for group bookings

- Adjust **deposit policy strategy**

- Introduce **demand-based pricing** during **peak months**

- Encourage **direct and corporate bookings**

- Reduce **long lead-time cancellation risk** via confirmation workflows

This dashboard transforms raw booking data into strategic intelligence for executive decision-making.

3.Sector & Business Context

Hospitality Industry Overview

The hotel industry **operates** under:

- High seasonality

- Variable demand patterns

- High OTA dependency

- Revenue management complexity

- Significant cancellation exposure

Cancellations directly impact:

- **Revenue** realization

- **Operational** planning

- **Staffing**

- **Pricing** strategy

Current Industry Challenges

- High OTA cancellations

- Price-sensitive segments

- Dynamic demand shifts

- Revenue leakage due to cancellation

- Inefficient deposit structures

Why This Problem Was Chosen

Cancellation management is a **core profitability** driver in hospitality. Understanding patterns across:

- Customer segments

- Booking behavior

- Deposit strategy

- Seasonal demand

allows hotels to improve forecasting, revenue protection, and policy design.

4.Problem Statement & Objectives

Problem Statement

"High cancellation rates and unpredictable demand patterns negatively impact hotel revenue and operational planning. Hotel management needs a data-driven way to understand booking behavior, identify cancellation drivers, and monitor revenue-related trends across hotel types, seasons, and customer segments".

Objectives

1. Measure cancellation rate across **segments**

2. **Identify** high-risk booking categories

3. **Analyze** seasonal revenue patterns

4. **Evaluate** deposit effectiveness

5. **Compare** repeat vs new guest stability

6. **Build** an interactive dashboard for decision-makers

Success Criteria

- Fully **functional interactive dashboard**

- Clear KPI mapping

- Actionable business recommendations

- Alignment between insights and strategy

---

5. Data Description

Dataset Source

Hotel Booking Demand Dataset (2015–2017)

Dataset Link :
https://www.kaggle.com/code/parameswarans1/hotel-basic-analysis-and-booking-demand-prediction

Size

- **~119,000+ records, 32 + columns**

- Multiple categorical & numerical features

Key Columns Used (for kpis, dashboard, pivot tables)

- Hotel

- Market Segment

- Distribution Channel

- Arrival Date (Year, Month)

- Lead Time

- Deposit Type

- is_cancelled

- is_repeated_guest

- ADR (Average Daily Rate)

Data Structure

Structured tabular dataset with categorical segmentation and numeric measures.

Limitations

- No external macroeconomic variables

- Limited behavioral history

- Deposit type anomaly suspected (even after selecting random data)

---

6. Data Cleaning & Preparation

All cleaning performed in **Google Sheets** as per capstone requirement.

Cleaning Steps

6.1 Removed blank records

- **Numeric fields** (e.g., children, babies): Missing values were imputed with 0.

- **Categorical fields** (e.g., country): Missing values were replaced with "Unknown".

- **Textual NULLs:** Standardized and treated as missing before imputation.

6.2 Outlier Treatment

Numeric variables such as ADR and days_in_waiting_list showed right-skewed distributions with some extreme values.

6.3 Transformations and Standardization

- **Data types:** Numeric columns were converted from text to numeric to ensure accurate calculations.

- **Text cleaning:** Categorical fields were standardized by trimming whitespace and normalizing labels.

- **Country mapping:** ISO country codes were mapped to full country names to improve dashboard readability.

6.4 Column Removal with Justification

Only variables relevant to cancellations, revenue, seasonality, and customer behavior were retained. Redundant, overly granular, or technical identifier columns were removed to simplify analysis and improve dashboard clarity.

6.5 Feature Engineering

The following derived features were created to enhance analysis:

- **Total Guests:** adults + children + babies

- **Total Stay Length:** weekday nights + weekend nights

- **Family Flag:** Family vs Non-Family bookings

- **Month Number:** Numeric month for chronological sorting

- **Revenue (Derived):** ADR × total stay length

6.6 Final Dataset Structure

The final, analysis-ready dataset includes:

- **Booking & Time:** hotel, arrival_date_year, arrival_date_month, lead_time

- **Customer & Channel:** market_segment, customer_type, deposit_type, distribution_channel, country
- **Behavior Metrics:** total_guests, total_stay_length, family_flag

- **Revenue Metrics:** ADR, derived revenue

- **Outcome:** is_canceled

---

7. KPI & Metric Framework

KPI Name	Definition	Formula	Business Importance	Objective Mapping
Total Bookings	Total confirmed booking records in the dataset	COUNT(booking_id)	Measures overall demand volume and booking inflow	Assess booking performance
Total Cancellations	Total number of cancelled bookings	SUM(is_cancelled)	Quantifies revenue leakage due to cancellations	Identify cancellation impact
Cancellation Rate (%)	Percentage of bookings that were cancelled	$\frac{\text{Total Cancellations}}{\text{Total Bookings}} \times 100$	Core risk indicator; reflects booking instability	Analyze cancellation behavior

Total Revenue	Total revenue generated from all bookings	SUM(Revenue)	Measures realized business performance	Evaluate financial performance
Average Revenue per Booking	Average revenue generated per booking	$\frac{\text{Total Revenue}}{\text{Total Bookings}}$	Indicates booking value and pricing efficiency	Improve pricing strategy

8. Exploratory Data Analysis (EDA)

This section integrates exploratory analysis with deeper evaluation to identify revenue trends, cancellation drivers, behavioral risk patterns, and strategic intervention points.

8.1 Revenue Trend & Seasonality

Monthly Revenue Trend Analysis

- Peak performance → **Month 7–8**

- Lowest performance → **Month 1 & 12**

- Booking volume follows identical pattern

Insight:
 Revenue variability is seasonal and demand-driven.

Business Impact:

- Apply dynamic pricing during peak periods

- Stimulate low-demand months via promotions

- Align operational capacity with seasonal cycles

8.2 Cancellation Risk Drivers

Market Segment Comparison

- Groups → ~54%

- Online TA → ~36%

- Corporate → ~16%

- Overall rate → 32.6%

Finding:
 Segment type significantly affects booking stability. Group bookings present the highest volatility.

Strategic Implication:
 Segment-specific policies are required to manage risk exposure.

Lead Time Distribution

- 180+ days → ~53%

- 91–180 days → ~38%

- 0–30 days → ~17%

Finding:
 Cancellation probability increases with booking horizon.

Strategic Implication:
 Long lead-time bookings require reconfirmation or stricter payment structures.

Deposit Policy Assessment

- Non-Refund → ~98%

- Refundable → 0%

- No Deposit → ~28%

Finding:
 The Non-Refund category shows abnormal behavior.

Risk / Anomaly Insight:
 Possible policy misclassification or enforcement gap.

Strategic Implication:
 Deposit structure requires audit and validation.

8.3 Behavioral & Stability Patterns

- Repeat guests cancel significantly less

- OTA-heavy segments show elevated volatility

- Short lead-time bookings demonstrate stronger commitment

Correlation Insight:
 Loyalty reduces risk. Early bookings increase uncertainty.

8.4 Advanced Evaluation

Root Cause Identification

Cancellation exposure is primarily driven by:

- Group bookings

- Long lead-time reservations

- OTA dependency

- Deposit inconsistencies

These structural drivers explain the overall 32.6% cancellation rate.

Risk Profiling

High-Risk Segments:

- Group bookings

- 180+ day bookings

- OTA channels

- Non-Refund deposit anomaly

Low-Risk Segments:

- Corporate bookings

- Repeat guests

- Short lead-time reservations

This enables targeted risk management rather than blanket policy changes.

Forecasting & Scenario Perspective

- Seasonality supports predictable revenue planning

- Lead-time distribution improves cancellation forecasting

- Reducing high-risk segment cancellations by even 5–10% would materially improve revenue realization

Small structural adjustments can create measurable financial gains.

8.5 Integrated Insight

Revenue volatility is not random.
 It is behavior-driven and structurally linked to:

- Segment type

- Booking timing

- Deposit structure

- Loyalty status

- Seasonal demand

The analysis establishes a clear foundation for strategic pricing, policy reform, risk control, and revenue optimization.

9. Dashboard Design

The dashboard is designed as a structured decision-support interface that converts booking data into actionable insights. It prioritizes clarity, risk visibility, and executive readability over visual complexity.

The layout follows a logical analytical progression:

- **Performance Snapshot:** Core KPIs summarize revenue, volume, and cancellation exposure.

- **Revenue Intelligence:** Seasonal trends highlight demand concentration and volatility.

- **Risk Diagnostics:** Segment behavior, lead time patterns, and deposit impact explain cancellation drivers.

- **Customer Stability:** Repeat guest behavior and channel performance reveal reliability patterns.

The design ensures that users can move from overall performance to root-cause analysis in a single view, enabling strategic evaluation without navigating multiple reports.

All analytical components operate within a unified filter-controlled environment, allowing dynamic assessment across key business dimensions.

Link : 📄 Dashboard Objective

10. Insights Summary

The following insights summarize the most critical findings derived from the analytical evaluation:

1. **Overall cancellation exposure is materially high (32.6%)**, representing a significant revenue realization risk.

2. **Group bookings present the highest instability (~54%)**, requiring immediate policy intervention.

3. **OTA channels generate volume but amplify cancellation risk (~36%)**, indicating a trade-off between scale and stability.

4. **Corporate and Direct bookings demonstrate superior reliability (~16%)**, making them strategically valuable segments.

5. **Long lead-time reservations (180+ days) show elevated cancellation probability (~53%)**, reducing forecast accuracy.

6. **Short lead-time bookings (0–30 days) reflect stronger customer commitment (~17%)**, contributing to booking certainty.

7. **Revenue is heavily seasonally concentrated**, with peak performance in mid-year months, reinforcing the need for demand-based pricing.

8. **Deposit policy effectiveness is questionable**, particularly within the Non-Refund category (~98% cancellation), indicating structural inconsistency.

9. **Customer loyalty materially reduces cancellation behavior**, as repeat guests demonstrate higher booking stability.

10. **Cancellation risk is behavior-driven and structurally predictable**, enabling targeted strategic intervention rather than blanket policy adjustments.

Pivot Wise **Insights** & detailed **Recommendations** : 📄 Insights

11. Key-Recommendations

Insight	Recommendation	Business Impact
1. High group cancellations	Stricter booking policy	Reduced revenue leakage
2. Long lead-time risk	Confirmation reminders	Improved booking stability
3. Seasonality	Demand-based pricing	Revenue optimization
4. Deposit anomaly	Policy audit	Risk correction
5. OTA risk	Incentivize Direct bookings	Margin improvement

12. Impact Estimation

Cost Savings

- **5% reduction** in the 32.6% cancellation rate → **direct revenue recovery**

- Lower booking reprocessing overhead

- **Reduced exposure** from high-risk segments (Group, OTA)

Operational Efficiency

- Lead-time controls improve occupancy forecasting

- Seasonal insights support staffing and inventory alignment

- Structured confirmation reduces volatility

Service Improvement

- Loyalty strengthening improves booking stability

- Direct & Corporate emphasis enhances customer control

- Clear deposit enforcement improves transactional reliability

Risk Reduction

- Targeted intervention in high-cancellation segments

- Policy correction for deposit anomaly

- Improved revenue predictability

---

13. Limitations

- Revenue estimated via **ADR proxy**

- **Non-Refund anomaly** may reflect data classification issue

- **No competitor** benchmarking included

- No **macroeconomic variables incorporated**

- **Descriptive analysis**; not predictive modeling

---

14. Future Scope

Advanced Analysis

- **Predictive** cancellation modeling.

- Time-series **revenue forecasting.**

- **Dynamic pricing simulation.**

Additional Data

- **Competitor** pricing data.

- **Customer lifetime** value metrics.

- **Real-time booking integration.**

These enhancements would enable predictive and prescriptive decision-making.

---

15. Conclusion

This project successfully transforms structured booking data into an executive-level analytical system.

The dashboard:

- **Identifies high-risk booking segments**

- **Quantifies cancellation exposure (32.6%)**

- **Reveals seasonal revenue concentration**

- **Evaluates lead-time and deposit policy effectiveness**

- **Highlights loyalty-driven booking stability**

- **Enables dynamic, filter-based strategic evaluation**

The analysis confirms that revenue volatility is structurally driven by **customer behavior, booking timing, channel mix, and policy design — not random variation.**

By **aligning pricing, deposit enforcement, and segment strategy** with these insights, the organization can move from reactive cancellation management to **proactive revenue optimization.**

18. Contribution Matrix

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Kasula Lalithendra	✅	✅	✅	✅	✅	✅	✅
Abhiman Singh	✅	✅			✅	✅	✅
Anant Singh	✅			✅		✅	✅
Rudraksh sharma	✅			✅			✅
Vridhi Chaudhary	✅		✅	✅		✅	✅
Ritik Raj	✅		✅			✅	✅

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

Team signature Block:

- Kasula Lalithendra
- Abhiman SIngh
- Vridhi Chaudhary
- Ritik Raj
- Anant Singh
- Rudraksh Sharma