



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Hotel Booking Intelligence Pipeline

The domain of the Project: Hotel Revenue and Customer Behavior Analytics

COURSE NAME: Data Analysis and Data Science

Team Mentors (and their designation):

Purnangshu Roy

AI Consultant at CSR BOX

Team Members:

Ms. Ajanta Ghosh

Period of the project

November 2025 to December 2025



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Declaration

The project titled “**Hotel Booking Intelligence Pipeline**” has been mentored by **Purnangshu Roy**, organised by **SURE Trust**, from **June 2025** to **December 2025**, for the benefit of the educated unemployed rural youth for gaining hands-on experience in working on industry relevant projects that would take them closer to the prospective employer. I declare that to the best of my knowledge the members of the team mentioned below, have worked on it successfully and enhanced their practical knowledge in the domain.

Team Members:

Ms. Ajanta Ghosh

Mentor's Name

AI Consultant—CSR BOX

Prof. Radhakumari

Executive Director & Founder

SURE Trust



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Table of contents

1. Executive summary
2. Introduction
3. Project Objectives
4. Methodology & Results
5. Social / Industry relevance of the project
6. Learning & Reflection
7. Future Scope & Conclusion



Executive Summary

The ***Hotel Booking Intelligence Pipeline*** is an end-to-end data analytics project designed to convert raw hotel reservation data into meaningful insights and predictive intelligence. The project integrates multiple technologies—**Excel**, **Python**—to analyze booking behavior, cancellation patterns, and revenue performance.

Initially, raw booking data was cleaned and transformed using Excel, where duplicates, missing values, and inconsistencies were handled, and new derived features were created. Python was used to build machine learning models for predicting booking status, estimating cancellation, and forecasting revenue, repeated guest prediction.

The outcomes of this project help hotel management identify high-risk bookings, understand factors influencing cancellations, and repeated guest analysis. Overall, the system demonstrates how data-driven intelligence can support strategic decision-making, improve operational efficiency, and enhance revenue optimization in the hospitality domain.



Introduction

Background and Context

The hotel industry generates vast amounts of data from bookings, customer preferences, and cancellations. Efficient analysis of this data can help hotels optimize revenue, improve occupancy, and enhance customer satisfaction. With the growth of online reservations, leveraging data-driven insights has become crucial for strategic decision-making.

Problem Statement

Hotels often face challenges such as last-minute cancellations, uneven occupancy, and inefficient pricing strategies. Manual analysis of booking data is time-consuming and prone to errors. There is a need for an automated system that can process booking data, predict cancellations, identify high-value customers, and provide actionable insights to optimize operations.

Scope and Limitations

Scope: The project analyzes hotel booking data to identify trends, predict cancellations, and highlight high-value customers. It provides interactive dashboards and insights to help hotels optimize revenue, improve occupancy, and enhance customer satisfaction.

Limitations: The analysis is limited by the quality and completeness of the dataset. External factors like seasonal events or sudden market changes may affect predictions. Real-time data integration and external market variables are not included.

Innovation Component

The project introduces an automated data pipeline that cleans, analyzes, and visualizes hotel booking data. It leverages predictive analytics to forecast cancellations and identify high-value bookings. Additionally, an interactive dashboard provides actionable insights for revenue optimization, improved occupancy management, and better customer experience.



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Project Objectives

The objectives of the project are as follows:

- To clean and preprocess raw hotel booking data for analysis
- To perform exploratory data analysis to identify key trends and patterns
- To analyze factors contributing to booking cancellations
- To build a predictive model to classify bookings as canceled or not canceled
- To demonstrate an end-to-end data science workflow using industry tools



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Project Objectives

Data Cleaning & Preparation: Ensure booking data is accurate, complete, and ready for analysis.

Trend Analysis: Identify patterns in bookings, cancellations, and customer behavior.

Predictive Modeling: Forecast cancellations and detect high-value bookings and repeated guest analysis using machine learning.

Decision Support: Enable data-driven strategies to enhance customer satisfaction and operational efficiency.



Methodology and Results

Methods / Technology Used:

Data Cleaning & Transformation: Handle missing values, duplicates, and outliers.

Data Analysis: Explore trends in bookings, cancellations, and customer behavior.

Predictive Modeling: Use machine learning models (e.g., Logistic Regression) to forecast cancellations and repeated guest bookings.

Visualization & Dashboarding: Represent insights using interactive charts and slicers.

Tools / Software Used:

Excel / Power Query: Data cleaning, transformation, and initial analysis.

Python (Pandas, Scikit-learn, Matplotlib, Seaborn): Predictive modeling and visualizations.

Data Collection Approach:

Historical hotel booking dataset collected from online hotel reservation records.

Includes attributes such as booking dates, lead time, number of guests, and cancellations.

Data was pre-processed to remove duplicates, handle missing values, and calculate derived metrics like net revenue and stay duration.

Project Architecture:

Data Input: Raw booking dataset from Excel/CSV.

Data Preprocessing: Cleaning, missing value handling, and feature engineering.

Analysis & Modeling: Trend analysis, predictive modeling for cancellations and high-value bookings and repeated guest booking analysis.

Visualization & Dashboard: charts, slicers, and reports for decision support.



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

Results:

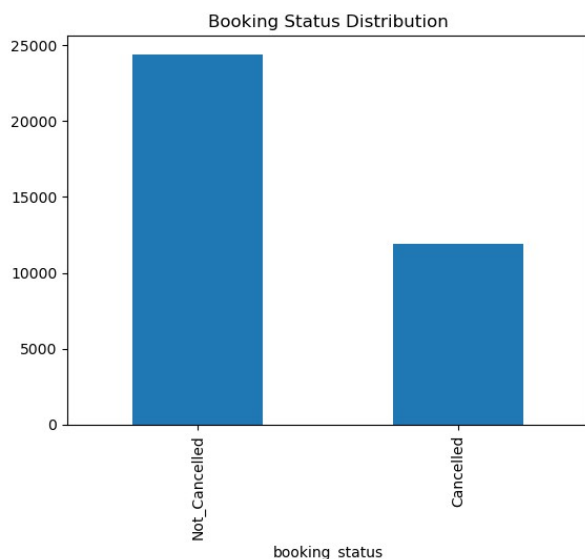
Predictive model accurately forecasts booking cancellations and highlights high-value customers.

Interactive dashboard allows filtering by repeated guest and booking status to monitor performance.

Insights help in revenue optimization by identifying high-value bookings and understanding booking trends.

Project GitHub Link:

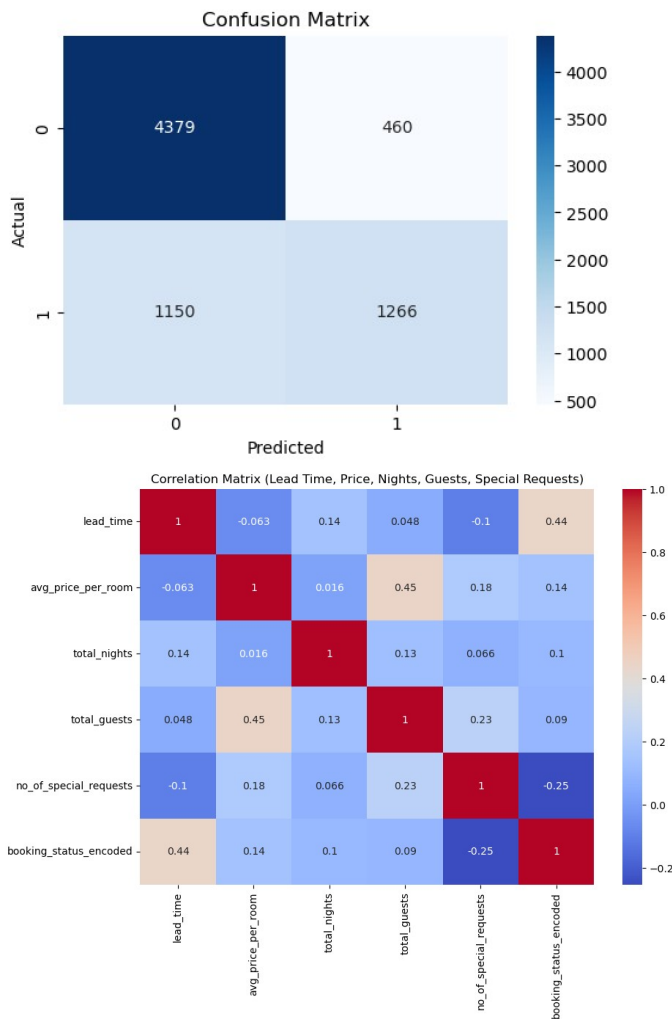
<https://github.com/ajantaGhost/HotelBookingProject>



A	B	C	D	E	F	G	H	I	J	K	L
Booking_ID	no_of_adults	no_of_children	no_of_weekend_nights	no_of_week_nights	type_of_meal_plan	required_car_parking_space	room_type_reserved	lead_time	arrival_year	arrival_month	arrival_date
INN00001	2	0	1	2	Meal Plan 1	0	Room_Type 1	224	2017	10	
INN00002	2	0	2	3	Not Selected	0	Room_Type 1	5	2018	11	
INN00003	1	0	2	1	Meal Plan 1	0	Room_Type 1	1	2018	2	
INN00004	2	0	0	2	Meal Plan 1	0	Room_Type 1	211	2018	5	
INN00005	2	0	1	1	Not Selected	0	Room_Type 1	48	2018	4	
INN00006	2	0	0	2	Meal Plan 2	0	Room_Type 1	346	2018	9	
INN00007	2	0	1	3	Meal Plan 1	0	Room_Type 1	34	2017	10	
INN00008	2	0	1	3	Meal Plan 1	0	Room_Type 4	83	2018	12	
INN00009	3	0	0	4	Meal Plan 1	0	Room_Type 1	121	2018	7	
INN00010	2	0	0	5	Meal Plan 1	0	Room_Type 4	44	2018	10	
INN00011	1	0	1	0	Not Selected	0	Room_Type 1	0	2018	9	
INN00012	1	0	2	1	Meal Plan 1	0	Room_Type 4	35	2018	4	
INN00013	2	0	2	1	Not Selected	0	Room_Type 1	30	2018	11	
INN00014	1	0	2	0	Meal Plan 1	0	Room_Type 1	95	2018	11	
INN00015	2	0	0	2	Meal Plan 1	0	Room_Type 1	47	2017	10	
INN00016	2	0	0	2	Meal Plan 2	0	Room_Type 1	256	2018	6	
INN00017	1	0	1	0	Meal Plan 1	0	Room_Type 1	0	2017	10	
INN00018	2	0	1	3	Not Selected	0	Room_Type 1	1	2017	8	
INN00019	2	0	2	2	Meal Plan 1	0	Room_Type 1	99	2017	10	
INN00020	2	0	1	0	Meal Plan 1	0	Room_Type 1	12	2017	10	
INN00021	2	0	2	2	Meal Plan 1	0	Room_Type 1	99	2017	10	
INN00022	1	0	0	1	Meal Plan 1	0	Room_Type 1	122	2018	11	
INN00023	2	0	2	4	Meal Plan 1	0	Room_Type 1	2	2018	3	
INN00024	2	0	0	3	Meal Plan 1	0	Room_Type 1	37	2018	10	
INN00025	2	0	2	1	Not Selected	0	Room_Type 1	130	2018	5	
INN00026	2	0	0	2	Meal Plan 1	0	Room_Type 1	99	2018	4	
INN00027	2	0	1	1	Meal Plan 1	0	Room_Type 1	60	2017	9	
INN00028	1	0	0	2	Meal Plan 1	0	Room_Type 4	2	2018	5	
INN00029	1	0	1	2	Meal Plan 1	0	Room_Type 1	37	2017	11	
INN00030	2	0	0	2	Meal Plan 2	0	Room_Type 1	56	2017	9	



Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)



Sum of booking id according to booking status		avg lead time according to status		ROOM TYPE	
Row Labels	Count of Booking_ID	Row Labels	Average of lead_time	Row Labels	Count of Booking_ID
Cancelled	11885	Cancelled	139.2154817	Room_Type 1	28130
Not_Cancelled	24390	Not_Cancelled	58.92722427	Cancelled	9072
Grand Total	36275	Grand Total	85.23255686	Not_Cancelled	19058
Sum of avg price per room		MARKET SEGMENT TYPE FILTER BY BOOKING STATUS		Room_Type 2	692
Row Labels	Average of avg_price_per_room	booking_status	Not_Cancelled	Cancelled	228
Cancelled	110.570467	Row Labels	Count of Booking_ID	Not_Cancelled	464
Not_Cancelled	99.92636326	Aviation	88	Room_Type 3	7
Grand Total	103.413756	Complementary	391	Cancelled	2
sum of guest according to status		Corporate	1797	Not_Cancelled	5
Row Labels	Sum of repeated_guest	Offline	7375	Room_Type 4	6057
Cancelled	16	Online	14739	Cancelled	2069
Not_Cancelled	914	Grand Total	24390	Not_Cancelled	3988
Grand Total	930	TOTAL NIGHTS		Room_Type 5	265
SPECIAL REQUEST		Row Labels	Average of total_nights	Cancelled	72
Row Labels	Sum of no_of_special_requests	Cancelled	3.279764409	Not_Cancelled	193
Cancelled	3977	Not_Cancelled	2.88601886	Room_Type 6	966
Not_Cancelled	18501	Grand Total	3.015024121	Cancelled	406
				Not_Cancelled	560
				Room_Type 7	158
				Cancelled	36
				Not_Cancelled	122
				Grand Total	36275



Learning and Reflection

Technical Learnings:

Gained hands-on experience in data cleaning, preprocessing, and feature engineering.

Learned to implement predictive models using Python and evaluate their performance.

Developed skills in data visualization.

Understood integration of multiple tools (Excel, Python) for end-to-end data analysis.

Management & Professional Learnings :

Improved project planning and time management to meet milestones.

Learned to document processes and present insights in a clear, professional manner.

Gained experience in working with large datasets and translating technical results into business insights.

Developed problem-solving and decision-making skills in a real-world scenario.

Overall Experience :

The project provided a comprehensive understanding of hotel data analytics from raw data to actionable insights.

Strengthened both technical and professional competencies.

Enhanced confidence in using data-driven approaches for operational and strategic decisions.

Overall, it was a valuable learning experience bridging theory and practical application.



Conclusion and Future Scope

Conclusion

The “Hotel Booking Intelligence Pipeline” successfully transformed raw booking data into actionable insights. By analyzing trends, predicting cancellations, and identifying high-value customers, the project demonstrated how data-driven decision-making can optimize revenue, improve occupancy, and enhance customer satisfaction. The interactive dashboard enables hotel management to quickly interpret data and implement strategies effectively.

Future Scope

Integrate **real-time data** for dynamic prediction and monitoring.

Enhance predictive models with **advanced machine learning techniques** (e.g., XGBoost, Neural Networks).

Include **external factors** like seasonal events, promotions, or competitor pricing for more accurate forecasting.

Expand the dashboard to provide **automated alerts and recommendations** for management decisions.

Explore **customer segmentation and personalized marketing strategies** to further increase revenue.