

Data Science Canvas				Project:	Hotel Booking Cancellation Prediction		
				Team:	Amit, Ghanshyam, Nitish, Shalin		
Problem Statement				Execution & Evaluation		Data Collection & Preparation	
Business Case & Value Added Hotel booking cancellations significantly impact revenue, occupancy forecasting, and resource planning. With flexible booking options and online travel platforms, cancellations have become frequent and unpredictable, leading to lost revenue and inefficient room utilization.	Model Selection 1. Logistic Regression 2. Random Forest 3. Gradient Boosting 4. XGBoost	Model Requirements Supervised classification, Simple interpretable models, For higher accuracy and non-linear relationships, avoid overfitting, Optimize model performance, Reliable performance estimation	Skills <ul style="list-style-type: none"> Python knowledge, Panda, Numpy, Scikit learn matplotlib, seaborn Understanding of Data cleaning, EDA concepts Knowledge of different regression and classification algorithms. 	Model Evaluation ROC-AUC – 0.9613 PR-AUC – 0.9429 Accuracy – 0.8938 Precision – 0.8712 Recall – 0.8370 F1-Score – 0.8538	Data Storytelling Hotel Management - High-level summary, visuals, dashboards Marketing & Sales Teams - Charts, heatmaps, and actionable recommendations Operations Team - Simple dashboards with daily/weekly cancellation predictions Data Science / IT Team - Technical report, model performance graphs, SHAP feature explanations	Data Selection & Cleansing <ul style="list-style-type: none"> Relevant features: Feature engineering to be done post EDA analysis Cleaning required: Yes — we'll clean for missing data, outliers, and possibly drop or transform features with low utility. 	Data Collection <ul style="list-style-type: none"> No separate data collection activity. Data will be collected from standard hotel booking and cancellation data sources <p>Not in scope of project work. We are using ready-made dataset from Kaggle.</p>
Data Landscape Dataset used : Hotel Booking Demand Dataset from Kaggle. Source: https://www.kaggle.com/datasets/mojtaba142/hotel-booking		Software & Libraries <ul style="list-style-type: none"> pandas, numpy – data handling, scikit-learn – modeling and evaluation matplotlib, seaborn – visualization xgboost – advanced boosting models 				Data Integration Not applicable for our project.	Explorative Data Analysis Detailed Univariate, Bivariate and Multivariate analysis done. <ul style="list-style-type: none"> - Encoding done for categorical features. - Scaling done for numeric features. - Added few more features as part of Feature Engineering. - Handled missing values for 4 features.