

**Word count:** 2000 +/-10% words; **Due Date:** Check Canvas; **Weight:** 20% (20 marks); **Submission:** Word file + R script in a zip file

## 1 Overview

Artificial Intelligence techniques can be applied to identify the key factors that drive business outcomes. In this assignment, you will use AI and machine learning methods to analyze hotel booking data and identify the most important business drivers that influence booking cancellations and revenue. You will apply feature selection techniques, quantify business impact, and develop actionable recommendations for hotel management.

## 2 Your Task

Your task is to analyze the Hotel Booking Demand dataset to identify key business drivers. You will apply the feature selection and business driver analysis techniques discussed in class. Your analysis must distinguish between actionable factors (that hotels can control) and non-actionable factors (external or fixed). You will quantify the financial impact of your findings and provide prioritized recommendations with ROI estimates.

You need to submit two files: (1) a Word document containing your business report and (2) an R script file with your complete analysis code. The report should be written for hotel executives who are not data scientists.

## 3 Dataset

**Dataset:** Hotel Booking Demand

**Download Link:** <https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand>

**File:** hotel.bookings.csv (119,390 records, 32 variables)

The dataset contains booking information for a city hotel and a resort hotel in Portugal from 2015-2017. Key variables include booking characteristics, guest information, distribution channels, and financial metrics. Your primary targets for analysis are booking cancellations (is\_canceled) and average daily rate (adr).

## 4 Assignment Structure

Your analysis should include:

### 4.1 Part A: Data Preparation (4 marks)

Load and explore the dataset. Handle missing values appropriately with business justification. Create visualizations to understand key patterns and distributions.

### 4.2 Part B: Business Driver Identification (8 marks)

Apply multiple methods to identify key drivers: correlation analysis, regression modeling, and feature importance techniques. Compare results across methods and categorize drivers as actionable or non-actionable.

### 4.3 Part C: Business Impact Analysis (5 marks)

Calculate the financial impact of key drivers. Perform what-if scenarios for realistic interventions. Estimate ROI for proposed changes.

### 4.4 Part D: Recommendations (3 marks)

Synthesize findings into specific, prioritized recommendations. Address implementation considerations and potential risks.

## 5 Submission Requirements

### 5.1 Files to Submit

1. **submit the following in a zip file:**
2. **Business Report:** Word document, 2000 words
3. **R Script:** Complete analysis code with comments

### 5.2 File Naming Convention

- Report: StudentID\_LastName\_A2\_Report.docx
- Code: StudentID\_LastName\_A2\_Code.R

## 6 Assessment Rubric

**Total Marks: 20** (Data Preparation: 4 + Business Driver Identification: 8 + Business Impact Analysis: 5 + Recommendations: 3)

Grade Level	Data Preparation (4 marks)	Business Driver Identification (8 marks)	Business Impact Analysis (5 marks)	Recommendations & Communication (3 marks)
<b>Fail (0-49%)</b>	Dataset not loaded OR code produces errors preventing execution. No data exploration visible.	No analysis methods completed OR code fails to produce results. Cannot identify any business drivers.	No financial calculations attempted OR calculations contain fundamental errors. No scenarios analyzed.	No clear recommendations OR report missing. Exceeds word limit by $\geq 20\%$ or under by $\geq 30\%$ . R code undocumented.
<b>Pass (50-59%)</b>	Dataset loaded successfully. Identifies 0-2 variables with missing values. Produces 1-2 basic plots (histogram or bar chart). Summary statistics for fewer than 3 variables.	Implements 1 method (correlation analysis only). Identifies 3-5 potential drivers without statistical support. Does not distinguish actionable vs non-actionable factors. No comparison of results.	Calculates basic financial impact for 1-2 drivers (e.g., revenue loss from cancellations). Attempts 1 what-if scenario but calculations incomplete. No ROI estimates provided.	Provides 2-3 vague recommendations (e.g., "improve booking process"). Report structure unclear with technical jargon unexplained. Within word limit $\pm 20\%$ . Basic code comments present.
<b>Credit (60-69%)</b>	Dataset loaded and explores 5+ variables. Handles missing values using one method (deletion OR mean imputation). Creates 3-4 meaningful visualizations. Calculates summary statistics for 5+ variables.	Implements 2 methods correctly (correlation + one other). Identifies 5-8 drivers with basic statistical support (p-values OR importance scores). Attempts to categorize actionable vs non-actionable with 2-3 examples. Shows results in tables.	Calculates financial impact for 3-4 drivers with clear formulas shown. Analyzes 2-3 what-if scenarios (e.g., "reduce cancellations by 10%"). Provides ROI estimate for 1 intervention with basic assumptions stated.	Provides 4-5 specific recommendations with basic prioritization (high/medium/low). Report follows logical structure with most technical terms explained. Creates 1-2 executive-appropriate visuals. Within word limit $\pm 10\%$ . Code sections clearly marked.
<b>Distinction (70-79%)</b>	Explores 10+ variables systematically. Handles missing values using 2+ methods with written justification (e.g., deletion for some, imputation for others). Creates 5-6 visualizations that reveal patterns. Documents data quality issues found.	Implements 3 methods correctly with comparison table. Identifies 8-12 drivers with clear statistical support (p-values AND effect sizes). Categorizes all drivers as actionable or non-actionable with justification. Ranks drivers by importance using consistent criteria.	Quantifies financial impact for 5+ drivers with detailed calculations. Analyzes 4-5 realistic scenarios with sensitivity analysis. Calculates ROI for 3+ interventions with break-even analysis. Identifies 2+ quick wins (high impact, low cost).	Provides 6-8 specific, measurable recommendations prioritized by impact and feasibility matrix. Executive summary highlights top 3 actions. All technical concepts explained in business terms. 3-4 professional visualizations included. Within word limit. Code fully documented with purpose of each section.

<b>High Distinction (80-100%)</b>	Explores all relevant variables with business context. Handles missing values using appropriate methods with clear business rationale (e.g., "Company ID missing likely means direct booking"). Creates 7+ professional visualizations with insights annotated. Identifies and addresses 3+ data quality issues.	Implements 3+ methods with sophisticated comparison (e.g., ensemble approach). Identifies 10+ drivers with comprehensive statistical evidence. Creates clear framework for actionable (e.g., "hotel controls directly") vs non-actionable (e.g., "market conditions"). Validates findings using hold-out data or cross-validation. Shows convergence/ divergence across methods.	Comprehensive financial model covering all major drivers. Analyzes 6+ scenarios with Monte Carlo simulation or confidence intervals. Provides ROI framework with payback periods, NPV calculations, and risk assessment. Creates implementation timeline showing quick wins vs strategic initiatives (6-12 months). Includes competitive benchmarking.	Provides 8+ SMART recommendations (Specific, Measurable, Achievable, Relevant, Time-bound) with detailed implementation roadmap. Executive dashboard with KPIs. Technical insights translated into compelling business narrative. 5+ boardroom-ready visualizations. Perfect word count. Code is reproducible with README instructions. Includes limitations and risk mitigation strategies.
-----------------------------------	--	--	--	--