Strategic Marketing Decision Making 2025-2026

Master in Data Science and Marketing Analytics
Erasmus School of Economics
Dr. Vardan Avagyan

Group Assignment 3
Due on: 7 October 2025, 11:00am

Instructions
This assignment is based on real hotel customer data. When completing this assignment, follow the step-by-step approach demonstrated in the lecture and textbook examples in Chapter 2. You may use the same model structures, data transformations, and R functions provided there. This will help you focus on interpretation rather than coding from scratch.

For data description, refer to Antonio, de Almeida, and Nunes (2020). To keep the analysis manageable, a random sample of 10,001 customers is provided. Since the dataset is based on real-world information, it requires cleaning and preparation before modelling. Remove customers with missing or non-realistic age values (i.e., "NULL", under 18, or over 95).[1]

Assume you work in the Data Science Department of the hotel. Management has approached your team to conduct a customer segmentation analysis, and to provide insights for targeting and positioning strategies. For running k-means cluster analysis, please use the seed "123" to ensure similarity of output across groups.

Question 1 (2 points)
What is the optimal number of clusters? Use the following variables: Age, Average Lead Time, Days Since Creation, Lodging Revenue, Other Revenue, Persons Nights, Room Nights, Days Since Last Stay, Days Since First Stay. Provide relevant figures to support your answer.

Question 2 (1 point)
Run a k-means cluster analysis using the number of clusters identified in Question 1. Create bar plots to show the cluster sizes and the overall segmentation solution.

Question 3 (2 point)
The hotel management aims to increase revenue and needs help identifying which customer segments to target. Based on your cluster analysis:

- Which cluster generates the highest lodging revenue?
- Which cluster contributes the most to other revenues (e.g., food, beverages, spa)?

Support your answer with bar plots and short justifications.

Question 4 (4 points)

[1] Note that an alternative approach is to impute missing age values, which is not needed for this assignment. Also, no additional data cleaning is required.

Management also requests guidance on how to position the hotel's services for the selected target segment(s). Propose two specific, actionable recommendations that are clearly grounded in your cluster analysis.

Appendix
Include your full R code in the appendix to ensure reproducibility of your results.