

**BUSINESS CASES FOR DATA SCIENCE**

**MASTER DEGREE PROGRAM IN DATA SCIENCE AND ADVANCED ANALYTICS**

BUSINESS CASE 2

Hotel booking demand

GROUP I

Emanuele Aldera, m20210617

Robin Schmidt, m20210602

Muhammad Abdullah, m20210655

Rui Ramos, m20210640

2nd March, 2022

INDEX

[1. Introduction iii](#_Toc98357614)

[2. Business understanding iii](#_Toc98357615)

[3. Data Preparation iii](#_Toc98357616)

[3.1 Data Transformation iii](#_Toc98357617)

[3.2 Outlier Treatment iii](#_Toc98357618)

[3.3 Data Normalization iii](#_Toc98357619)

[3.4 Subsetting iii](#_Toc98357620)

[4. Clustering iii](#_Toc98357621)

[4.1 K-Means iii](#_Toc98357622)

[5. Results of the Clustering iii](#_Toc98357623)

[6. Business Approach iv](#_Toc98357624)

[7. Appendix v](#_Toc98357625)

# Introduction

For this business case we receive data describing hotel demand. We been able to access to data with observations related to a city hotel.

# Business understanding

The main goal of this kind of segments is to understand bookings cancellation prediction, but also important in this case make some segmentation of customers since we have two different approaches taking in account the different positioning in the two different kinds of hotels. Also the seasonality could be an important factor to give a good perspective of the strategy to be taken trough different seasons.

Our goal in this case is the prediction of booking cancelations and the reason behind.

# Data Understanding

In the datasets data don’t present missing values although we find 32% of duplicates most of them in “Type” variables group all of them categorical and fully manually registered by the staff (I’m assuming guys need to check with you this)

Some interesting conclusions:

* ReservationDayOfWeek shows higher cancelation rates during the week
* 1208 null values(~5%)in ADR
* Room P is always canceled

# Data Preparation

## 3.1 Data Transformation

We decide to transform some variables to get a simplified and coherent dataset, so we build new variables, like duration (sum of weekend and week nights), we created a new one “Is Room changed”, and transformed arrival date to date format, and Reservation Date from datetime format to individual date features.

## 3.2 Outlier Treatment

## 3.3 Data Normalization

## 3.4 Subsetting

# Clustering

## 4.1 K-Means

# Results of the Clustering

# Business Approach

# Appendix

Figure 1

Figure 2

Gráfico, Gráfico de barras

Descrição gerada automaticamente

Figure 3

Gráfico

Descrição gerada automaticamente

Figure 4

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Figure 5

Gráfico

Descrição gerada automaticamente

Figure 6

Figure 7

Figure 8

Figure 9

Figure 10

Figure 11a

Figure 11b

Figure 11c

Figure 11d

Figure 12

Figure 13a

Figure 13b

Figure 13c

Figure 13d

Figure 14

Figure 15

Figure 16

Figure 17

Figure 18

Figure 19

Figure 20

Figure 21