

**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 ii](#_Toc98450063)

[2. Business understanding iii](#_Toc98450064)

[3. Data Understanding iii](#_Toc98450065)

[3.1. Data Preparation iii](#_Toc98450066)

[3.2. Duplicate data iii](#_Toc98450067)

[3.3. Missing values iv](#_Toc98450068)

[4. Data Exploration iv](#_Toc98450069)

[5. Data Transformation iv](#_Toc98450070)

[6. Appendix v](#_Toc98450071)

# 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 and predict bookings cancellation, but also important in this case, to make some segmentation of customers since we have an approach taking in account the rotation of this kind of hotels. 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 dataset 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)

## Data Preparation

* 30 variables 🡪 Discrete, Categorical and Continuous
* In order to simplify we start to do some feature Engineering, from 9 variables related with arrivals and stays we transform 9 variables to 6 ("DurationOfStay", "TotalValue", "ArrivalDateMonthNumber", "ArrivalDate", "ArrivalDate", 'ArrivalDayOfWeek') and drop one ("ArrivalDateMonth"), this one was duplicated since we have get the arrival date month number.
* Some arrangements was made to date type in order to get more readable data.
* After we have the categorical and numerical data treated we include all in the dataset

## Duplicate data

* We find aprox. 32% of duplicated data, some treatment was required so that we could avoid overfiting. In this case we decided to keep one of the values so that we could keep the most of data possible. After treatment we get 86% of records kept after removal.

## Missing values

* Was detected in children and country some missing values, they where in small account so we change this Nan values by the mode.

Some interesting conclusions:

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

# Data Exploration

* In terms of cardinality of the target variable is obvious the not canceled wins ([Fig.1](#_Data_Exploration))
* Visual analysis with Histograms of metric variables show us unbalanced data (Fig.2)
* By market segment we could analyze by far bigger cardinality in online TA

## Correlations

* Was chosen spearman correlation then linear one, looking at the scatter where more adequate because major of correlations are linear

# Data Transformation

# Appendix

Figure 1

Gráfico, Gráfico de barras

Descrição gerada automaticamente

Figure 2

Gráfico, Histograma

Descrição gerada automaticamente

Figure 3

Gráfico, Gráfico de barras

Descrição gerada automaticamente

Figure 4

Interface gráfica do usuário

Descrição gerada automaticamente

Gráfico, Histograma

Descrição gerada automaticamente

Gráfico, Gráfico de barras

Descrição gerada automaticamente

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Interface gráfica do usuário, Gráfico

Descrição gerada automaticamente

Gráfico, Gráfico de cascata

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