

Práctica 2: Booking Data Cleaning

Gerard Carrera and Guillem Rochina

2022-12-24

DESCRIPCIÓN DEL DATASET

```
booking <- read.csv("hotels_data.csv")
summary(booking)
```

```
##      name              city      check.in      adults
## Length:2632      Length:2632      Length:2632      Min.   :2.00
## Class :character      Class :character      Class :character      1st Qu.:2.00
## Mode  :character      Mode  :character      Mode  :character      Median :4.00
##                                           Mean   :3.32
##                                           3rd Qu.:5.00
##                                           Max.   :5.00
##
##      children      check.out      num_rooms      address
## Min.   :1.000      Length:2632      Min.   :2.000      Length:2632
## 1st Qu.:1.000      Class :character      1st Qu.:2.000      Class :character
## Median :3.000      Mode  :character      Median :2.000      Mode  :character
## Mean   :2.484
## 3rd Qu.:3.000
## Max.   :3.000
##
##      hotel_coordinates      hotel_score      hotel_scores      hotel_description
## Length:2632      Min.   :-1.00      Length:2632      Length:2632
## Class :character      1st Qu.: 8.00      Class :character      Class :character
## Mode  :character      Median : 8.40      Mode  :character      Mode  :character
##                                           Mean   : 8.25
##                                           3rd Qu.: 8.80
##                                           Max.   :10.00
##                                           NA's   :5
##      features      room_data      page_count      current_page
## Length:2632      Length:2632      Min.   : 0.00      Min.   :1.000
## Class :character      Class :character      1st Qu.: 25.00      1st Qu.:2.000
## Mode  :character      Mode  :character      Median : 54.00      Median :3.000
##                                           Mean   : 57.53      Mean   :2.758
##                                           3rd Qu.: 88.00      3rd Qu.:4.000
##                                           Max.   :133.00      Max.   :5.000
##
##      in_page_count      search_date
## Min.   : 0.00      Length:2632
## 1st Qu.: 6.00      Class :character
```

```
## Median :13.00   Mode  :character
## Mean    :12.55
## 3rd Qu. :19.00
## Max.    :27.00
##
```

INTEGRACIÓN Y SELECCIÓN

```
##                                name            city
## 1 Holiday Inn Express Barcelona City 22@, an IHG Hotel Barcelona
## 2                                Ofelias Hotel 4* Sup Barcelona
## 3                                Onix Fira Barcelona
## 4                                Aspasios Rambla Catalunya Suites Barcelona
## 5                                Hotel Lloret Ramblas Barcelona
##      check.in adults children      check.out num_rooms
## 1 12-December-2022      2      3 16-December-2022      2
## 2 12-December-2022      2      3 16-December-2022      2
## 3 12-December-2022      2      3 16-December-2022      2
## 4 12-December-2022      2      3 16-December-2022      2
## 5 12-December-2022      2      3 16-December-2022      2
##                                address
## 1      Pallars, 203, Sant Mart\xed, 08005 Barcelona, Spain
## 2      Llan\xed, 24, Eixample, 08015 Barcelona, Spain
## 3      Llan\xed, 30, Eixample, 08015 Barcelona, Spain
## 4      Rambla de Catalunya 42, Eixample, 08007 Barcelona, Spain
## 5 Rambla Canaletes, 125, Ciutat Vella, 08002 Barcelona, Spain
##      hotel_coordinates hotel_score
## 1 41.40145712,2.19775379      8.1
## 2 41.37673208,2.15001449      8.6
## 3 41.37705814,2.14974493      7.7
## 4 41.39020152,2.16497585      8.6
## 5 41.38493925,2.17013776      8.1
##
## 1 {'': '', 'Staff': '9.0', 'Facilities': '8.0', 'Cleanliness': '8.4', 'Comfort': '8.2', 'Value for m
## 2 {'': '', 'Staff': '9.0', 'Facilities': '8.7', 'Cleanliness': '9.2', 'Comfort': '9.1', 'Value for m
## 3 {'': '', 'Staff': '8.7', 'Facilities': '7.5', 'Cleanliness': '8.1', 'Comfort': '7.9', 'Value for m
## 4 {'': '', 'Staff': '8.3', 'Facilities': '8.6', 'Cleanliness': '8.8', 'Comfort': '8.9', 'Value for m
## 5 {'': '', 'Staff': '8.9', 'Facilities': '7.8', 'Cleanliness': '8.4', 'Comfort': '8.2', 'Value for m
##
## 1
## 2 ["You're eligible for a Genius discount at Ofelias Hotel 4* Sup! To save at this property, all you
## 3
## 4
## 5
##
## 1 ['Free WiFi', 'Bathtub', 'Air conditioning', '24-hour front desk', 'Daily housekeeping', 'Non-smok
## 2
## 3
## 4      ['Apartments', 'Kitchen', 'Washing machine', 'Free WiFi', 'Air conditioning', 'Private Bathroom
## 5
##
## 1
## 2 {'35087504_359192576_0_2_0': {'room_name': 'Triple Room (3 Adults)', 'room_price': '\x80 723', 'ro
```

```
## 3
## 4
## 5
##  page_count current_page in_page_count search_date
## 1          0           1             0 2022-11-18
## 2          1           1             1 2022-11-18
## 3          2           1             2 2022-11-18
## 4          3           1             3 2022-11-18
## 5          4           1             4 2022-11-18
```

LIMPIEZA DE LOS DATOS

ANÁLISIS DE LOS DATOS

RESOLUCIÓN DEL PROBLEMA