Universidad Peruana de Ciencias Aplicadas



**ANÁLISIS EXPLORATORIO DE UN CONJUNTO DE DATOS EN R/RSTUDIO**

Mauro Imanol Obermeyer Adauto Angulo - 202111197

David Joaquín Niño Suárez

Jair Stephano Rodríguez Calla

Facultad de Ingeniería, Ciencias de la Computación

Fundamentos de Data Science – CC51

Patricia Daniela Reyes Silva

28 de abril de 2023

**ÍNDICE**

CONTENIDO

[1. Caso de análisis 2](#_Toc133602419)

[2. Conjunto de datos (Data Set) 2](#_Toc133602420)

[3. Análisis de datos exploratorio 2](#_Toc133602421)

[a. Carga de datos 2](#_Toc133602422)

[b. Inspección de datos 2](#_Toc133602423)

[c. Visualización gráfica 2](#_Toc133602424)

[4. Conclusiones preliminares 2](#_Toc133602425)

**CONTENIDO**

## **Caso de análisis**

## **Conjunto de datos (Data Set)**

|  |  |
| --- | --- |
| Variable | Descripción |
| hotel | Resort Hotel o City Hotel |
| is\_canceled | Valor que indica si la reserva ha sido cancelada (1) o no (0) |
| lead\_time | Número de días que pasaron entre la fecha de la reservada agendada en el Sistema Operativo Hotelero y la fecha de llegada |
| arrival\_date\_year | Año de fecha de llegada |
| arrival\_date\_month | Mes de fecha de llegada |
| arrival\_date\_week\_number | Número de semana del año para la fecha de llegada |
| arrival\_date\_day\_of\_month | Día de fecha de llegada |
| stays\_in\_weekend\_nights | Número de noches de fin de semana (sábado o domingo) que el huésped se hospedó o agendó hospedaje en el hotel |
| stays\_in\_week\_nights | Número de noches de semana (lunes a viernes) que el huésped se hospedó o agendó hospedaje en el hotel |
| adults | Número de adultos |
| children | Número de niños |
| babies | Número de bebés |
| meal | Tipo de comida reservada. Undefined/SC – paquete sin comida; BB – Bed & and Breakfast - Desayuno y cama; HB – Half Board - Medio paquete; FB – Full Board - Paquete completo |
| country | País de origen (Formato ISO) |
| market\_segment | Mercado de segmento objetivo. TA – Travel Agents – Agentes de viaje; TO – Tour Operators – Operadores turísticos |
| distribution\_channel | Canal de distribución de reservas. TA – Travel Agents – Agentes de viaje; TO – Tour Operators – Operadores turísticos |
| is\_repeated\_guest | Valor que indica si el nombre en la reserva es de un huésped repetido (1) o no (0) |
| previous\_cancellations | Número de reservas anteriormente canceladas por un huésped antes de la presente reserva |
| previous\_bookings\_not\_canceled | Número de reservas anteriormente no canceladas por un huésped antes de la presente reserva |
| reserved\_room\_type | Código del tipo de habitación reservado. Presentado en vez de la designación por razones de anonimato. |
| assigned\_room\_type | Código del tipo de habitación asignado a la reserva. A veces el tipo de habitación asignado difiere del reservado por razones de funcionamiento del hotel. Presentado en vez de la designación por razones de anonimato. |
| booking\_changes | Número de cambios hechos en la reserva desde el momento en el que ha sido ingresada al Sistema Operativo Hotelero hasta el momento del check-in o cancelación. |
| deposit\_type | Indicación si el cliente ha hecho un depósito para garantizar la reserva. No Deposit – No Depósito – no se ha hecho un depósito; Non Refund – No Reembolso – depósito realizado en el valor del costo total de hospedaje; Refundable – Reembolsable – depósito realizado con valor menor al costo total de hospedaje |
| agent | ID de la agencia de viaje que hizo la reserva |
| company | ID de la compañía/entidad que hizo la reserva o responsable de pagar la reserva. Presentado en vez de la designación por razones de anonimato |
| days\_in\_waiting\_list | Número de días que la reserva ha estado en la lista de espera antes de ser confirmada por el cliente |
| customer\_type | Tipo de reserva, asumiendo una de cuatro categorías: Contract – Contrato – cuando la reserva tiene asociada una asignación u otro tipo de contrato; Group – Grupo – cuando la reserva está asociada a un grupo; Transient – Transitorio – cuando la reserva no es parte de un grupo o contrato, y no está asociada a otro tipo de reserva transitoria; Transient-party – cuando la reserva es transitoria, pero está asociada a, por lo menos, otra reserva transitoria |
| adr | Average Daily Rate (Ratio promedio diario), definido al dividir la suma de todas las transacciones de alojamiento entre el número total de noches de estadía. |
| required\_car\_parking\_spaces | Número de espacios de estacionamiento requeridos por el cliente |
| total\_of\_special\_requests | Número de pedidos espaciales hechos por el cliente |
| reservation\_status | Último estado de reservación. Canceled - Cancelado – la reserva fue cancelada por el cliente; Check-Out – Verificar – el cliente se registró, pero ya se ha ido; No-Show – No Mostrar – el cliente no hizo el chek-in e informó al hotel el motivo |
| reservation\_status\_date | Fecha en la que se colocó el último estado. Puede ser usada en junto a ReservationStatus para entender cuánto se canceló la reserva o cuándo se retiró el cliente |

## **Análisis de datos exploratorio**

### **Carga de datos**

Utilizando el siguiente comando leemos el data set (.csv, con coma (“,”) como separador).

hotel\_data\_NULL <- read.csv("ruta/hotel\_bookings.csv")

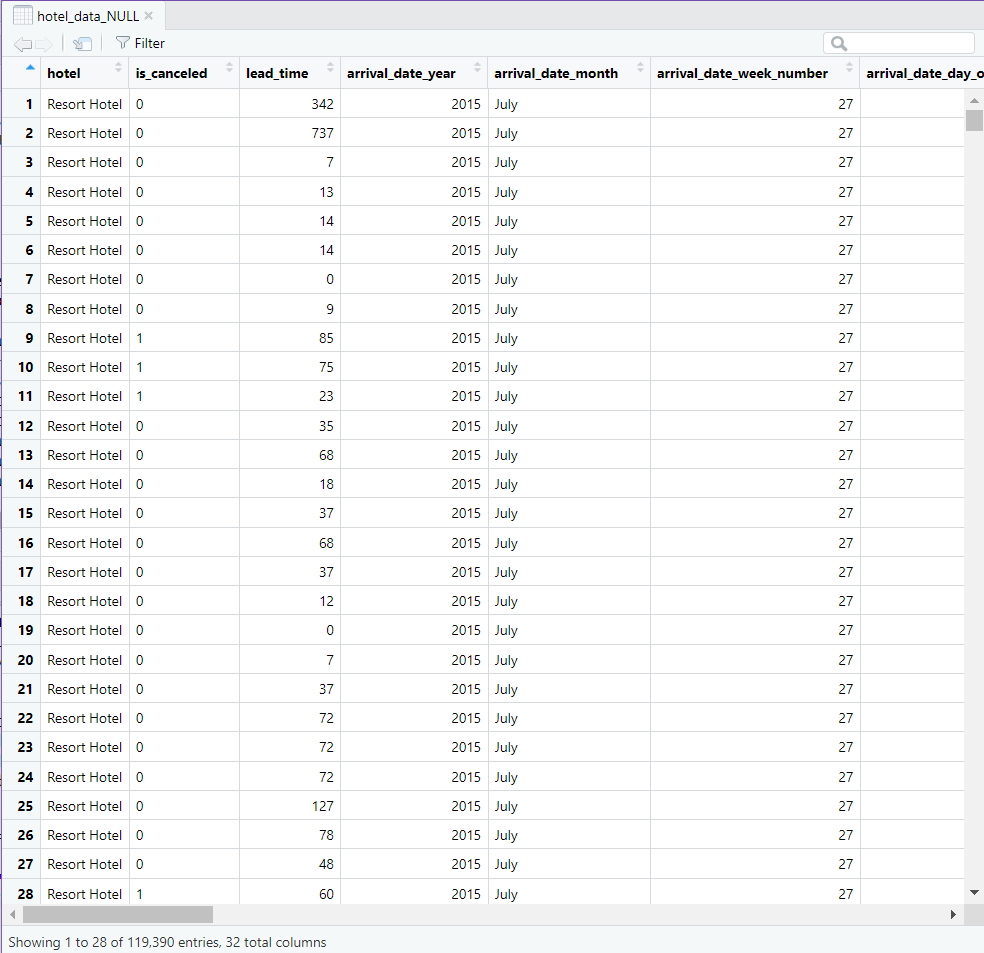
Con este comando también leemos el data set, pero reemplaza todos los datos NULL por *NA*.

hotel\_data\_NONULL <- read.csv("ruta/hotel\_bookings.csv", na.strings = "NULL")

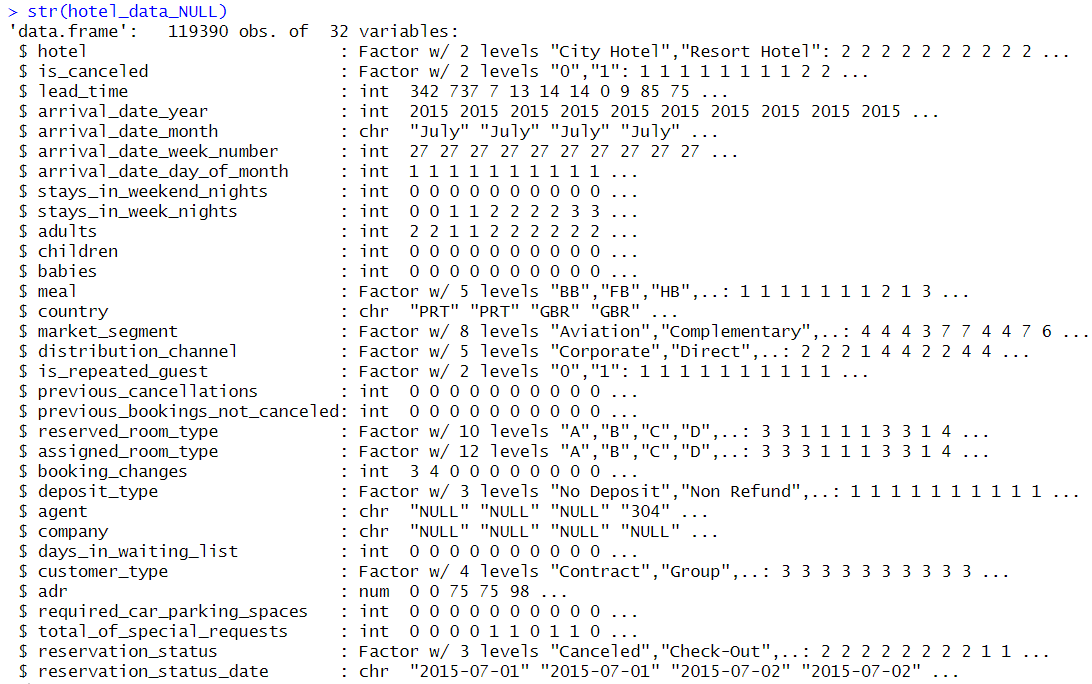
### **Inspección de datos**

Ejecutamos el siguiente comando View(hotel\_data\_NULL) para visualizar el data set mediante una tabla.

Cada columna de variables tiene los datos correctamente establecidos. Sin embargo, las columnas agent y company carecen de datos por lo que son reemplazados por NULL.

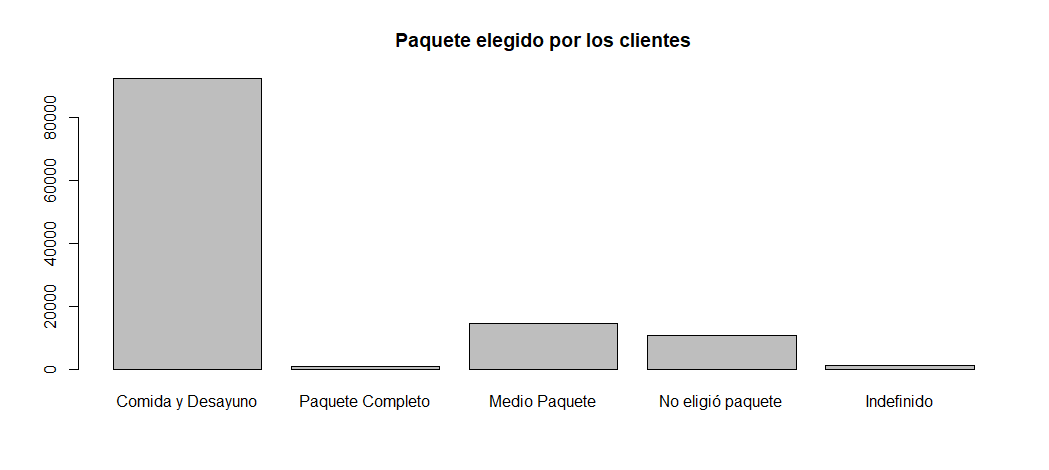


Mediante el comando str(hotel\_data\_NULL) visualizamos las distintas variables en la consola, donde podemos saber el tipo de dato de cada una. Gracias a esto, podemos volver factores a las variables que tengan un número de valores definido con el comando hotel\_data\_NULL$nombre\_variable <- as.factor(hotel\_data\_NULL$nombre\_variable).



### **Visualización gráfica**

* Gráfico de los paquetes más elegidos en las reservas



## **Conclusiones preliminares**