

Universidad Peruana de Ciencias Aplicadas (UPC)



Proyecto:

Anàlisis exploratorio de un conjunto de datos en R/Rstudio

Curso:

Administraciòn de la Informaciòn

Integrantes:

- Carhuanchu Bazan, Alessandro David u201913933
- Nuñez Lazo, Sergio Antonio u201910357
- Cruz Mamani, Jack Yefri u201912400

Ciclo:

2021-01

Indice

1. Caso de análisis
 - 1.1. Origen de los datos
 - 1.2. Casos de uso aplicables
2. Conjunto de datos
 - 2.1. Descripción de la estructura de datos
 - 2.2. Resumen de la estructura de datos
 - 2.3. Deteccion de valores NA
 - 2.4. Detección de valores vacios
 - 2.5. Deteccion de valores outliers
3. Análisis de datos exploratorios
 - 3.1. Cantidad de clientes hospedados en cada hotel por año
 - 3.2. Cantidad de reservas canceladas y no canceladas por año
 - 3.3. Distribución de clientes que vienen solo entre adultos y con menores de edad por hotel
4. Conclusiones preliminares
5. Referencias

1. Caso de análisis

1.1. Origen de los datos

El dataset analizado en el presente trabajo fue obtenido desde los repositorios de Kaggle, sin embargo, este dataset ya ha pasado por el proceso de limpieza de datos, el cual fue realizado Thomas Mock y Antoine Bichat para los repositorios de Tidy Tuesday en GitHub, un proyecto semanal enfocado en el análisis de datos usando el lenguaje de programación R. El commit fue realizado el 11 de Febrero del 2020.

El dataset original proviene del artículo Hotel Booking Demand Datasets. Según el artículo, el dataset fue extraído de las bases de datos Property Management System SQL de los dos hoteles examinados. Dichos hoteles se ubican en Portugal, precisamente en el Resort de Algarve, y en la Ciudad de Lisboa.

El artículo está disponible en los repositorios de Science Direct desde el 29 de noviembre del 2018. En el mes de Febrero del año 2019, el artículo fue publicado en el volumen 22 de la revista Data in Brief. Los autores son Nuno Antonio ^{a, b}, Ana Almeida ^{a, c, d}, y Luis Nunes ^{a, b, d}, quienes realizaron estudios en las siguientes instituciones de Portugal:

- a. Instituto Universitário de Lisboa, Lisboa
- b. Instituto de Telecomunicações, Lisboa
- c. CISUC, Coimbra
- d. ISTAR-IUL, Lisboa

1.2. Casos de uso aplicables

La utilidad del dato es innegable, sin embargo, la importancia de este dataset en particular se refleja en los siguientes campos:

- ☐ Análisis descriptivo para el entendimiento de patrones, tendencias y anomalías en la data.
- ☐ Investigación para la predicción de cancelación de una reserva y satisfacción del cliente, dicha información beneficia a los hoteles.
- ☐ Benchmarking de algoritmos de clasificación y segmentación, el cual beneficia a los investigadores y educadores de machine learning.
- ☐ Entrenamiento de modelos de data mining y estadística para beneficio de los educadores.

2. Conjunto de datos (DataSet)

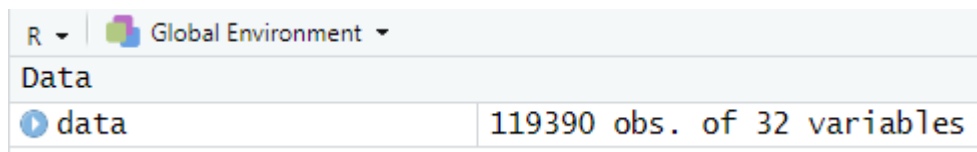
2.1. Descripción de la estructura de datos

Para la descripción de la estructura de datos se utilizaron los siguientes comandos:

- 1) Para leer el conjunto de datos del archivo .csv usamos la instrucción **read.csv()**.

```
> data <- read.csv("C:/Users/PC/Desktop/GitHub/TA1_Administracion/data/hotel_bookings.csv",  
+                 header=TRUE, sep=",")
```

Al lado izquierdo de Rstudio, se puede observar que presenta 119390 observaciones y 32 variables.



- 2) Usamos la instrucción **head()** para obtener las cinco primeras observaciones.

```
> head(data)
  hotel is_canceled lead_time arrival_date_year arrival_date_month arrival_date_week_number arrival_date_day_of_month stays_in_weekend_nights stays_in_week_nights adults children babies meal
1 Resort Hotel      0      342          2015           July          27              1              0              0      2      0      0      0 BB
2 Resort Hotel      0      737          2015           July          27              1              0              0      2      0      0      0 BB
3 Resort Hotel      0       7          2015           July          27              1              0              1      1      0      0      0 BB
4 Resort Hotel      0      13          2015           July          27              1              0              1      1      0      0      0 BB
5 Resort Hotel      0      14          2015           July          27              1              0              2      2      0      0      0 BB
6 Resort Hotel      0      14          2015           July          27              1              0              2      2      0      0      0 BB
  country market_segment distribution_channel is_repeated_guest previous_cancellations previous_bookings_not_canceled reserved_room_type assigned_room_type booking_changes deposit_type agent company
1 PRT Direct Direct 0 0 0 C C 3 No Deposit NULL NULL
2 PRT Direct Direct 0 0 0 C C 4 No Deposit NULL NULL
3 GBR Direct Direct 0 0 0 A C 0 No Deposit NULL NULL
4 GBR Corporate Corporate 0 0 0 A A 0 No Deposit 304 NULL
5 GBR Online TA TA/TO 0 0 0 A A 0 No Deposit 240 NULL
6 GBR Online TA TA/TO 0 0 0 A A 0 No Deposit 240 NULL
  days_in_waiting_list customer_type adr required_car_parking_spaces total_of_special_requests reservation_status reservation_status_date
1 0 Transient 0 0 0 Check-Out 2015-07-01
2 0 Transient 0 0 0 Check-Out 2015-07-01
3 0 Transient 75 0 0 Check-Out 2015-07-02
4 0 Transient 75 0 0 Check-Out 2015-07-02
5 0 Transient 98 0 1 Check-Out 2015-07-03
6 0 Transient 98 0 1 Check-Out 2015-07-03
```

- 3) Obtenemos los nombres de las columnas con la instrucción **names()**.

```
> names(data)
 [1] "hotel"
 [6] "arrival_date_week_number"
[11] "children"
[16] "distribution_channel"
[21] "assigned_room_type"
[26] "days_in_waiting_list"
[31] "reservation_status"

      "is_canceled"
      "arrival_date_day_of_month"
      "babies"
      "is_repeated_guest"
      "booking_changes"
      "customer_type"
      "reservation_status_date"

      "lead_time"
      "stays_in_weekend_nights"
      "meal"
      "previous_cancellations"
      "deposit_type"
      "adr"

      "arrival_date_year"
      "stays_in_week_nights"
      "country"
      "previous_bookings_not_canceled"
      "agent"
      "required_car_parking_spaces"

      "arrival_date_month"
      "adults"
      "market_segment"
      "reserved_room_type"
      "company"
      "total_of_special_requests"
```

- 4) Obtenemos el tipo de dato de cada variable y los valores que contiene con la instrucción **str()**.

```
> str(data)
'data.frame': 119390 obs. of 32 variables:
 $ hotel                : Factor w/ 2 levels "City Hotel","Resort Hotel": 2 2 2 2 2 2 2 2 2 ...
 $ is_canceled          : int 0 0 0 0 0 0 0 0 1 1 ...
 $ lead_time            : int 342 737 7 13 14 14 0 9 85 75 ...
 $ arrival_date_year    : int 2015 2015 2015 2015 2015 2015 2015 2015 2015 ...
 $ arrival_date_month   : Factor w/ 12 levels "April","August",...: 6 6 6 6 6 6 6 6 6 ...
 $ arrival_date_week_number : int 27 27 27 27 27 27 27 27 27 ...
 $ arrival_date_day_of_month : int 1 1 1 1 1 1 1 1 1 ...
 $ stays_in_weekend_nights : int 0 0 0 0 0 0 0 0 0 ...
 $ stays_in_week_nights  : int 0 0 1 1 2 2 2 2 3 3 ...
 $ adults               : int 2 2 1 1 2 2 2 2 2 ...
 $ children             : int 0 0 0 0 0 0 0 0 0 ...
 $ babies               : int 0 0 0 0 0 0 0 0 0 ...
 $ meal                 : Factor w/ 5 levels "BB","FB","HB",...: 1 1 1 1 1 1 1 2 1 3 ...
 $ country              : Factor w/ 178 levels "ABW","AGO","AIA",...: 137 137 60 60 60 60 137 137 137 ...
 $ market_segment      : Factor w/ 8 levels "Aviation","Complementary",...: 4 4 4 3 7 7 4 4 7 6 ...
 $ distribution_channel : Factor w/ 5 levels "Corporate","Direct",...: 2 2 2 1 4 4 2 2 4 4 ...
 $ is_repeated_guest    : int 0 0 0 0 0 0 0 0 0 ...
 $ previous_cancellations : int 0 0 0 0 0 0 0 0 0 ...
 $ previous_bookings_not_canceled : int 0 0 0 0 0 0 0 0 0 ...
 $ reserved_room_type   : Factor w/ 10 levels "A","B","C","D",...: 3 3 1 1 1 1 1 3 3 1 4 ...
 $ assigned_room_type   : Factor w/ 12 levels "A","B","C","D",...: 3 3 3 1 1 1 3 3 1 4 ...
 $ booking_changes      : int 3 4 0 0 0 0 0 0 0 ...
 $ deposit_type         : Factor w/ 3 levels "No Deposit","Non Refund",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ agent                : Factor w/ 334 levels "1","10","103",...: 334 334 334 157 103 103 334 156 103 40 ...
 $ company              : Factor w/ 353 levels "10","100","101",...: 353 353 353 353 353 353 353 353 353 ...
 $ days_in_waiting_list : int 0 0 0 0 0 0 0 0 0 ...
 $ customer_type        : Factor w/ 4 levels "Contract","Group",...: 3 3 3 3 3 3 3 3 3 3 ...
 $ adr                  : num 0 0 75 75 98 ...
 $ required_car_parking_spaces : int 0 0 0 0 0 0 0 0 0 ...
 $ total_of_special_requests : int 0 0 0 0 1 1 0 1 0 ...
 $ reservation_status   : Factor w/ 3 levels "Canceled","Check-Out",...: 2 2 2 2 2 2 2 2 1 1 ...
 $ reservation_status_date : Factor w/ 926 levels "2014-10-17","2014-11-18",...: 122 122 123 123 124 124 124 124 73 62 ...
```

5) Obtenemos un resumen de cada variable con la instrucción **summary()**. Esto nos devuelve, en caso de variables numéricas, los siguientes valores:

- el valor mínimo (Min)
- el primer cuartil (1st Qu)
- la mediana (Median)
- la media (Mean)
- el tercer cuartil (3rd Qu)
- el valor máximo (Max)

Por otro lado, para las variables de tipo String o categórica:

- tipo 1 : cantidad 1
- tipo 2 : cantidad 2
- tipo 3 : cantidad 3
- tipo n : cantidad n

```
> summary(data)
      hotel      is_canceled      lead_time      arrival_date_year      arrival_date_month      arrival_date_week_number      arrival_date_day_of_month      stays_in_weekend_nights      stays_in_week_nights
City Hotel :79330      Min. : 0.0000      Min. : 0      Min. :2015      August :13877      Min. : 1.00      Min. : 1.0      Min. : 0.0000      Min. : 0.0
Resort Hotel:40060      1st Qu.:0.0000      1st Qu.: 18      1st Qu.:2016      July   :12661      1st Qu.:16.00      1st Qu.: 8.0      1st Qu.: 0.0000      1st Qu.: 1.0
      Median :0.0000      Median : 69      Median :2016      May    :11791      Median :28.00      Median :16.0      Median : 1.0000      Median : 2.0
      Mean :0.3704      Mean :104      Mean :2016      October:11160      Mean :27.17      Mean :15.8      Mean : 0.9276      Mean : 2.5
      3rd Qu.:1.0000      3rd Qu.:160      3rd Qu.:2017      April  :11089      3rd Qu.:38.00      3rd Qu.:23.0      3rd Qu.: 2.0000      3rd Qu.: 3.0
      Max. :1.0000      Max. :737      Max. :2017      June   :10939      Max. :53.00      Max. :31.0      Max. :19.0000      Max. :50.0
      NA's :4
      adults      children      babies      meal      country      market_segment      distribution_channel      is_repeated_guest      previous_cancellations
Min. : 0.000      Min. : 0.0000      Min. : 0.000000      BB      :92310      PRT      :48590      Online TA :56477      Corporate:6677      Min. : 0.0000      Min. : 0.00000
1st Qu.: 2.000      1st Qu.: 0.0000      1st Qu.: 0.000000      FB      : 798      GBR      :12129      Offline TA:24219      Direct   :14645      1st Qu.:0.00000      1st Qu.: 0.00000
Median : 2.000      Median : 0.0000      Median : 0.000000      HB      :14463      FRA      :10415      Groups    :19811      GDS      : 193      Median :0.00000      Median : 0.00000
Mean : 1.856      Mean : 0.1039      Mean : 0.007949      SC      :10650      ESP      : 8568      Direct    :12606      TA/TO    :97870      Mean :0.03191      Mean : 0.08712
3rd Qu.: 2.000      3rd Qu.: 0.0000      3rd Qu.: 0.000000      Undefined:1169      DEU      : 7287      Corporate :5295      Undefined: 5      3rd Qu.:0.00000      3rd Qu.: 0.00000
Max. :55.000      Max. :10.0000      Max. :10.000000      (Other):28635      ITA      : 3766      Complementary:743      (Other): 239      Max. :1.00000      Max. :26.00000
      previous_bookings_not_canceled      reserved_room_type      assigned_room_type      booking_changes      deposit_type      agent      company      days_in_waiting_list      customer_type
Min. : 0.0000      A      :85994      A      :74053      Min. : 0.0000      No Deposit:104641      9      :31961      NULL      :112593      Min. : 0.000      Contract : 4076
1st Qu.: 0.0000      D      :19201      D      :25322      1st Qu.: 0.0000      Non Refund:14587      NULL      :16340      40      : 927      1st Qu.: 0.000      Group    : 577
Median : 0.0000      E      : 6535      E      : 7806      Median : 0.0000      Refundable: 162      240      :13922      223      : 784      Median : 0.000      Transient:89613
Mean : 0.1371      F      : 2897      F      : 3751      Mean : 0.2211      (Other):2211      1      : 7191      67      : 267      Mean : 2.321      Transient-Party:25124
3rd Qu.: 0.0000      G      : 2094      G      : 2553      3rd Qu.: 0.0000      (Other):42797      14      : 3640      45      : 250      3rd Qu.: 0.000
Max. :72.0000      B      :1118      C      : 2375      Max. :21.0000      (Other):4354      7      : 3539      153      : 215      Max. :391.000
      NA's :4
      adr      required_car_parking_spaces      total_of_special_requests      reservation_status      reservation_status_date
Min. : -6.38      Min. : 0.0000      Min. : 0.0000      Canceled :43017      2015-10-21: 1461
1st Qu.: 69.29      1st Qu.:0.00000      1st Qu.:0.0000      Check-Out:75166      2015-07-06: 805
Median : 94.58      Median :0.00000      Median :0.0000      No-Show : 1207      2016-11-25: 790
Mean :101.83      Mean :0.06252      Mean :0.5714      (Other) :114477      2015-01-01: 763
3rd Qu.:126.00      3rd Qu.:0.00000      3rd Qu.:1.0000      (Other) :114477      2016-01-18: 625
Max. :5400.00      Max. :8.00000      Max. :5.0000      (Other) :114477      2015-07-02: 469
```

2.2. Resumen de la estructura de datos

El dataset presenta 119390 observaciones (40060 resort hotel + 793 30 city hotel) y 32 variables. Sus variables se muestran en la siguiente tabla:

Nombre	Tipo de dato	Descripción
hotel	Factor	El tipo de hotel (city, resort)
is_canceled	Int	Si se ha cancelado (1), sino (0)
lead_time	Int	El tiempo entre la reserva y la llegada al hotel (días)
Arrival_date_year	Int	Año de llegada al hotel
Arrival_date_month	Int	Mes de llegada(12 categorías de “Enero” a “Diciembre”)
Arrival_data_week_number	Int	Número de semana de llegada al hotel
Arrival_date_day_of_month	Int	Día del mes de la llegada al hotel
Stays_in_weedend_nights	Int	Número de noches de fines de semanas reservados o prescenciados (“Sábado o Domingo”)
Stays_in_week_nights	Int	Número de noches de semana reservados o prescenciados (“Lunes” a “Viernes”)
Adults	Int	Número de adultos
Children	Int	Número de niños
Babies	Int	Número de bebés
Meal	Factor	Tipo de comida reservada: - NA/“SC”: Sin paquete de comida - “BB”: Bed & Breakfast - “HB”: Half board, solo desayuno, no cena
Country	Factor	País de origen (las categorías están representadas en ISO 3155-3:2013 format)
Market_segment	Factor	Segmento de mercado destinado - “TA”: agencia de viajes - “TO”: operaciones de tour
Distribution_channel	Factor	Medio por el cual se realizó la reserva: - “TA”: Travel agents - “TO”: Tour operations
Is_repeated_guest	Int	Si es usuario repetido (1), sino (0)
Previous_cancellations	Int	Número de reservaciones canceladas anterior a la reserva
Previous_bookings_not_canceled	Int	Número de reservaciones no canceladas anterior a la reserva
Reserved_room_type	Factor	Código de tipo de habitación reservado
Assigned_room_type	Factor	Código de tipo de habitación asignado (algunas veces no se le asigna la habitación reservada, ya sea por doblereserva o petición del usuario)

Booking_changes	Int	Número de cambios hechos a la reserva desde la fecha de reserva hasta el check-in.
Deposit_type	Factor	Si realizó deposito para garantizar la reserva: - “ No Deposit”: no se deposito - “Not refund”: se deposito el valor total o mayor de la reserva - “Refundable”: se deposito un valor menor al costo total
Agent	Factor	El Id de la agencia de viajes el cual realizo la reserva
Company	Factor	El Id de la compañía que realizó la reserva o responsable a quien se pago la reserva.
Days_in_waiting_list	Int	Número de días que la reserva estaba en lista antes de que fuera confirmada al cliente
Customer_type	Factor	Tipo de reserva: - Contract: La reserva es asociado a un contrato o persona quien funciona como intermediario - Group: La reserva es asociada a un grupo - Transient: No es parte de grup or contract, y no asociado a una reserva transitorio - Transient-party: Cuando ha sido asociado con al menos un transitorio.
Adr	Num	Promedio de ventas diarias
Requiered_car_parking_spaces	Int	Número de espacios de para estacionar el vehículo por el cliente
Total_of_special_rerquest	Int	Número de pedidos espciales por el cliente (dos camas, piso mayor, etc.)
Reservation_status	Factor	El último estado de la reservación: Cancelled: Fue cancelado por el cliente Check-out: Realizó el check-in, pero ya se fue. No Show: El cliente no realizó check-in pero informó al hotel los motivos
Reservation_status_date	Factor	Fecha en el cual el último Reservation_status fue establecido.

2.3. Detección de valores NA

Para la detección de valores NA, creamos una funcion que muestre solo solo valores NA de un dataframe.

```

> funcion.NA <- function(x){
+   sum = 0
+   for(i in 1:ncol(x)){
+     cntNA <- colSums(is.na(x[i]))
+     if(cntNA > 0){
+       cat("En la columna ", colnames(x[i]), " total de valore NA", cntNA, "\n")
+     }
+   }
+ }
>
> funcion.NA(data)
En la columna  children  total de valore NA 4

```

Obtenemos que la columna children presenta valores NA.

2.4. Detección de valores vacios

Para la detección de valores vacios creamos una funcion que nos muestre de todas las columnas, sus valores vacios.

```

> function.Vacios <- function(x){
+   for(i in 1:ncol(x)){
+     cat("En la columna",colnames(x[i]),"total de valores en blanco:",colSums(x[i]==""),"\n")
+   }
+ }
>
> function.Vacios(data)

```

Obtenemos el siguiente resultado:

```

En la columna hotel total de valores en blanco: 0
En la columna is_canceled total de valores en blanco: 0
En la columna lead_time total de valores en blanco: 0
En la columna arrival_date_year total de valores en blanco: 0
En la columna arrival_date_month total de valores en blanco: 0
En la columna arrival_date_week_number total de valores en blanco: 0
En la columna arrival_date_day_of_month total de valores en blanco: 0
En la columna stays_in_weekend_nights total de valores en blanco: 0
En la columna stays_in_week_nights total de valores en blanco: 0
En la columna adults total de valores en blanco: 0
En la columna children total de valores en blanco: NA
En la columna babies total de valores en blanco: 0
En la columna meal total de valores en blanco: 0
En la columna country total de valores en blanco: 0
En la columna market_segment total de valores en blanco: 0
En la columna distribution_channel total de valores en blanco: 0
En la columna is_repeated_guest total de valores en blanco: 0
En la columna previous_cancellations total de valores en blanco: 0
En la columna previous_bookings_not_canceled total de valores en blanco: 0
En la columna reserved_room_type total de valores en blanco: 0
En la columna assigned_room_type total de valores en blanco: 0
En la columna booking_changes total de valores en blanco: 0
En la columna deposit_type total de valores en blanco: 0
En la columna agent total de valores en blanco: 0
En la columna company total de valores en blanco: 0
En la columna days_in_waiting_list total de valores en blanco: 0
En la columna customer_type total de valores en blanco: 0
En la columna adr total de valores en blanco: 0
En la columna required_car_parking_spaces total de valores en blanco: 0
En la columna total_of_special_requests total de valores en blanco: 0
En la columna reservation_status total de valores en blanco: 0
En la columna reservation_status_date total de valores en blanco: 0
~|

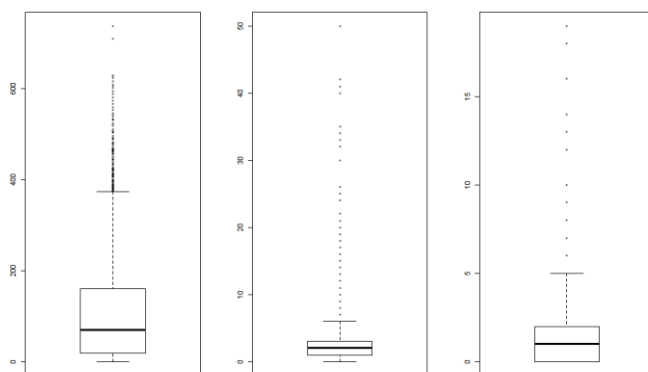
```


2.5. Detección de valores outliers

Para la detección de outliers utilizamos la función **boxplot()**, este modelo usa la detección de outliers por el método de las cajas.

```
> #outliers
> par(mfrow = c(1,3))
> boxplot(data$lead_time)
> boxplot(data$stays_in_week_nights)
> boxplot(data$stays_in_weekend_nights)
```

Obteniendo el siguiente resultado, donde los puntos representan los outliers:



3. Análisis de datos exploratorios

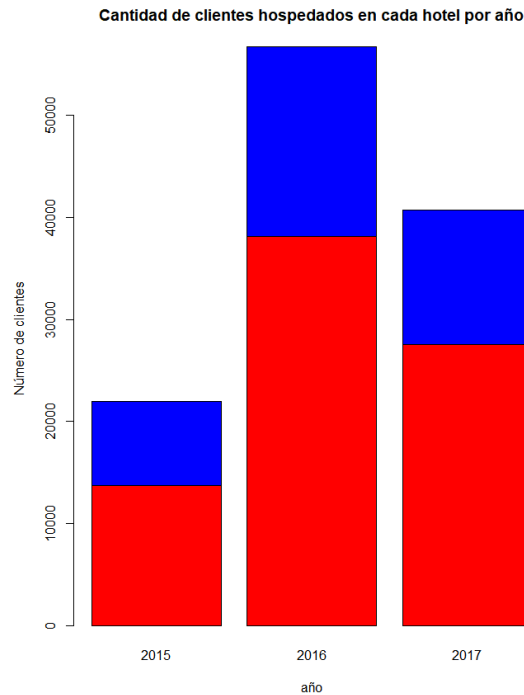
3.1. Cantidad de clientes hospedados en cada hotel por año

Para hallar la cantidad de clientes hospedados en cada hotel por año, utilizamos la instrucción **table()**, luego usamos **barplot()** para imprimirlo.

```
> ClientesPorHotel=table(data$hotel,data$arrival_date_year)

> barplot(ClientesPorHotel,
+         main="Cantidad de clientes hospedados en cada hotel por año",
+         col=c("red","blue"),legend=c("City hotel","Resort hotel"),
+         xlab = ("año"),ylab = ("Número de clientes"),args.legend = list(x = "topright",inset = c(- 0.05, -0.25)))
```

Obteniendo el siguiente resultado:



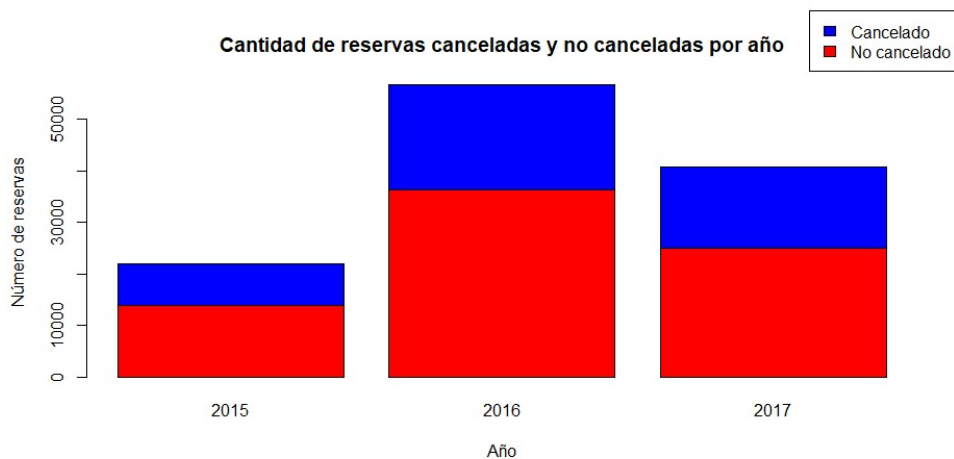
3.2. Cantidad de reservas canceladas y no canceladas por años

Para hallar la cantidad de reservas canceladas y no canceladas por años, utilizamos la instrucción **table()**, luego usamos **barplot()** para imprimirlo.

```
> canceladosPorAnio=table(data$is_cancelled,data$arrival_date_year)

> barplot(canceladosPorAnio,
+         main="Cantidad de reservas canceladas y no canceladas por año",
+         col=c("red","blue"),legend=c("no cancelado","cancelado"),
+         xlab = ("Año"),ylab = ("Número de reservas"),args.legend = list(x = "topright",inset = c(- 0.05, -0.25)))
```

Obteniendo el siguiente resultado:



3.3. Distribución de clientes que vienen solo entre adultos y con menores de edad por hotel

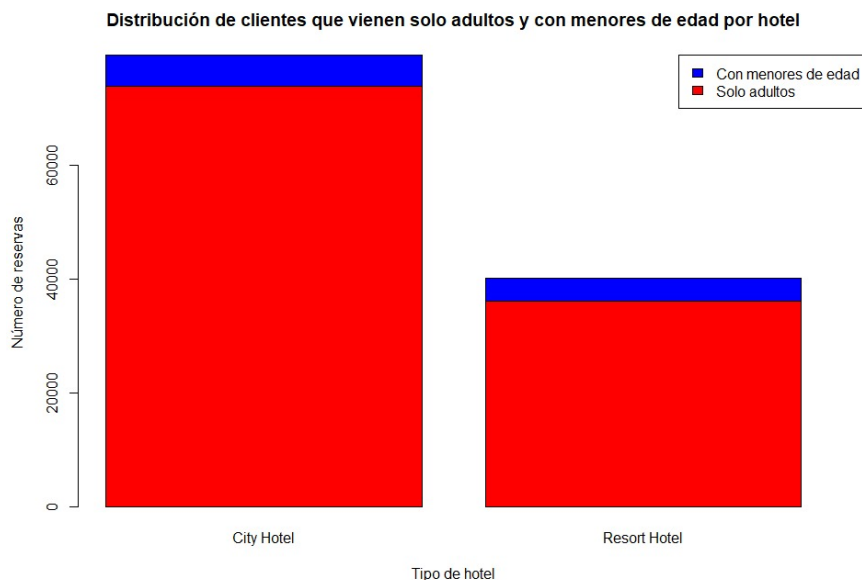
Para hallar la cantidad de clientes que vienen solo entre adultos y con menores de edad por hotel, sumamos los cantidad de reservas con bebés y con niños, luego, si el resultado es mayor a cero, entonces la reserva se realizó con menores, sino solo entre adultos. Finalmente, utilizamos la instrucción **table()** y **barplot()** para imprimirlo.

```
> #Tabla 3: Distribución de clientes que vienen solo entre adultos y con menores de edad por hotel
> menores <- data$babies + data$children > 0
> summary(menores)
  Mode FALSE  TRUE  NA's
logical 110054 9332    4

> menoresDeEdad = table(menores, data$hotel)

> barplot(menoresDeEdad,
+         main="Distribución de clientes que vienen solo adultos y con menores de edad por hotel",
+         col=c("red","blue"),legend=c("Solo adultos", "Con menores de edad"),
+         xlab = ("Tipo de hotel"),ylab = ("Número de reservas"),args.legend = list(x = "topright",inset = c(- 0.05, -0.25)))
~ |
```

Obteniendo el siguiente resultado:



4. Conclusiones preliminares

Dados el análisis de las tres tablas anteriores podemos deducir las siguientes conclusiones:

- 1) El hotel de tipo ciudad es más concurrido que el de tipo resort durante los 3 años.

- 2) El año 2016 fue el año con más turistas, pero a la vez fue el año con más cancelaciones, además, si el número de reservas crece, también lo hace las cancelaciones.
- 3) La mayor cantidad de personas viajan solos o en grupo, pero que entren en la categoría de adultos.

5. Referencias

Hotel booking demand datasets Antonio, de Almeida, & Nunes - Data in Brief – 2019. Recuperado de <https://doi.org/10.1016/j.dib.2018.11.126>

Link del GitHub https://github.com/u201912400/TA1_Administracion