

# **Sistema de Reservas de Hotel**

## **Informe Operaciones Preliminares**

Mayra Goicochea Neyra

## CONTENIDO

CONTENIDO	2
ABSTRACT	3
INTRODUCCION	3
OBJETIVOS	4
OPERACIONES COMUNES	4
OPERACIONES ESPECIFICAS A LA ASIGNATURA TECNICAS DE AGRUPACION Y REDUCCION DE LA DIMENSION	8
OPERACIONES ESPECIFICAS A LA ASIGNATURA PREDICCION	9
CONCLUSIONES	10
REFERENCIAS BIBLIOGRAFICAS	10
ANEXO	10

## **ABSTRACT**

Las cancelaciones de reservas son una parte inevitable de la industria hotelera, y en los últimos años han incrementado significativamente. Esto afecta directamente a la gestión de la demanda de los Administradores de hoteles y resorts, sobre todo en la optimización del inventario del hotel y la estrategia de revenue.

Algunos especialistas indican que las mayores tasas de cancelación se encuentran en los canales de venta online, debido a las facilidades a los usuarios que proporcionan algunas OTA's (Online Travel Agencies como Expedia, Booking, Hotels.com entre otras) que han provocado un comportamiento de reserva en varios hoteles y luego quedarse con solo uno.

En el presente informe, se explica la construcción de un prototipo basado en datos (anonimizados) de la demanda de un hotel tipo resort ubicado en la ciudad de Algarve (Portugal). El objetivo del prototipo es solucionar aspectos importantes de la problemática en la industria hotelera: la detección de qué reservas son "susceptibles de cancelar" (mediante un modelo predictivo), clasificación de los usuarios en grupos (basados en sus similitudes) para que el hotel pueda gestionar estrategias de fidelización o captación de clientes de forma óptima (estas agrupaciones son relacionadas a una tasa de cancelación para detectar algunos grupos hitos) y finalmente, generará el pronóstico del número de reservas semanales.

## **INTRODUCCIÓN**

En los últimos años, las cancelaciones de reservas en hoteles han incrementado significativamente (La tasa promedio en el 2014 fue de 32.9% y en el 2017 ha alcanzado al 41.3% [1]). Para reducir el impacto, los

administradores o encargados de la gestión revenue, implementan políticas de cancelación rígidas o estrategias de overbooking, que puede generar reducción de ingresos o dañar su imagen comercial. Las políticas de cancelación rígidas pueden percibirse con una reducción en el número de reservas (los usuarios encuentran más atractivas las ofertas de otros hoteles), y sobre el overbooking, puede generar malestar a los clientes cuando hay mayor demanda que las habitaciones disponibles.

En cuanto al impacto de las cancelaciones, los hoteles pierden la oportunidad de prestar un servicio a otros usuarios, y se incrementa el impacto cuanto más cerca sea la cancelación a la fecha de checkin.

Algunos estudios, muestran que el incremento de la tasa de cancelaciones se debe a la presencia de OTAs (Online Travel Agencies como Expedia, Booking, Hotels.com entre otras), las que promueven crear un sentimiento de urgencia para que los usuarios reserven a pesar de no haber tomado una decisión (con mensajes alusivos como "Alta Demanda", "Mejor Precio", "Última Habitación" entre otras) [2]. También afirman que las reservas con un tiempo de anticipación mayor a 60 días son 65% más probables de cancelar [1]. Con la estimación de que reservas son probables que se cancelen, los gerentes y administradores podrán efectuar medidas para evitarlo (como ofrecer servicios, mejoras, entre otros), además de gestionar logísticamente sus habitaciones y servicios. Paralelamente, con el prototipo, también se puede identificar grupos de clientes basados en características y así ofrecerles a ellos un servicio personalizado a sus atributos comunes, de esa forma se puede reducir la tasa de cancelaciones y entregar valor al negocio. Finalmente, el prototipo estima el número semanal de reservas para gestionar la oferta, y que los encargados de la administración puedan tomar decisiones de mediano y largo plazo como ampliación de servicios y/o habitaciones, campañas de publicidad y ofertas entre otras.

## OBJETIVOS

La industria hotelera es un sector muy competitivo, donde se tiene mucha oferta (compiten en precio o servicios adicionales entre los más principales) y cuya demanda es variable, pero también muestra comportamientos estacionales (periodos de alta demanda en temporadas de verano o festivos). Por ello es importante brindar al gerente del hotel mecanismos para alcanzar una buena posición en el mercado. Este prototipo generará más valor al negocio, entregando métricas e información consistente y precisa para la toma de decisiones ante las problemáticas identificadas previamente como:

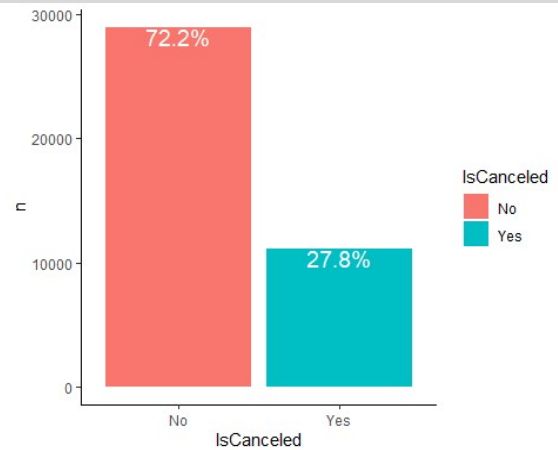
- Identificación de segmentos de usuarios para estrategias de marketing y venta.
- Detección de reservas muy probables de ser canceladas
- Pronóstico de la demanda semanal en reservas.

## OPERACIONES COMUNES

El conjunto de datos contiene 40060 observaciones de reservas, seleccionadas según su fecha de arribo, comprendidas en el periodo del 1 de julio del 2015 al 31 de agosto del 2017. Para iniciar la construcción de los modelos de clasificación y predictivos, se realizó un análisis exploratorio de los datos. Con el summary, no se presentan observaciones con valores ausentes.

La muestra se encuentra distribuida en 72.2% de Reservas no canceladas y 27.8% de Reservas canceladas (Ver Fig. 1).

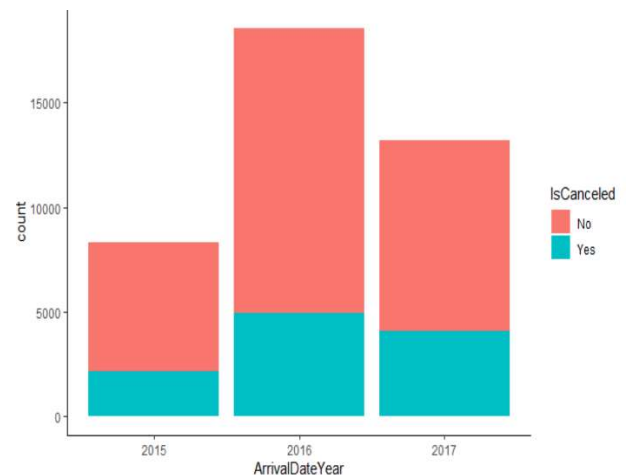
**Fig: 1.**



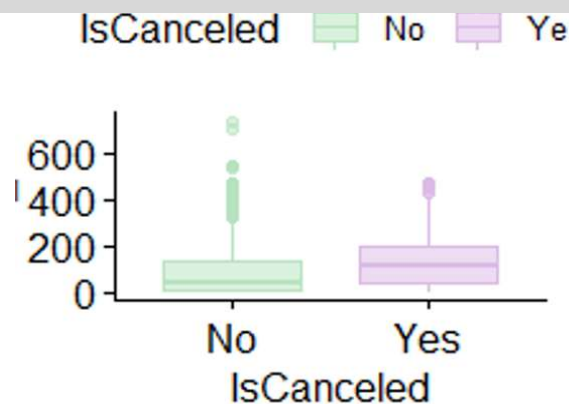
Las distribuciones anuales muestran similar proporción, siendo el año 2016, el que tiene mayor registro de reservas (Ver Fig. 2).

Cuando el valor de **Lead Time** aumenta se tiene mayor probabilidad de que se cancele la reserva, lo vemos en el gráfico de cajas, donde el rango de **Canceled** abarca aproximadamente desde Lead Time 100 a más (Ver Fig.3).

**Fig: 2.**

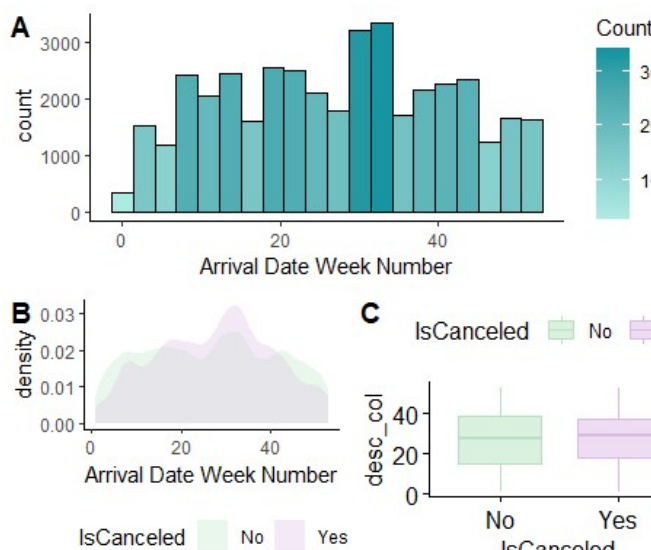


**Fig: 3.**



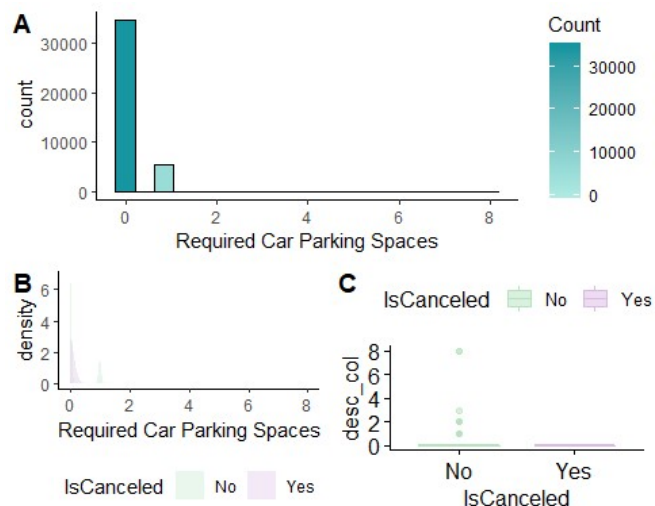
Las **semanas** 30 y 35 tienen mayor número de reservas, y la diferencia es significativa entre las canceladas y el caso contrario (Ver Fig.4).

**Fig: 4.**



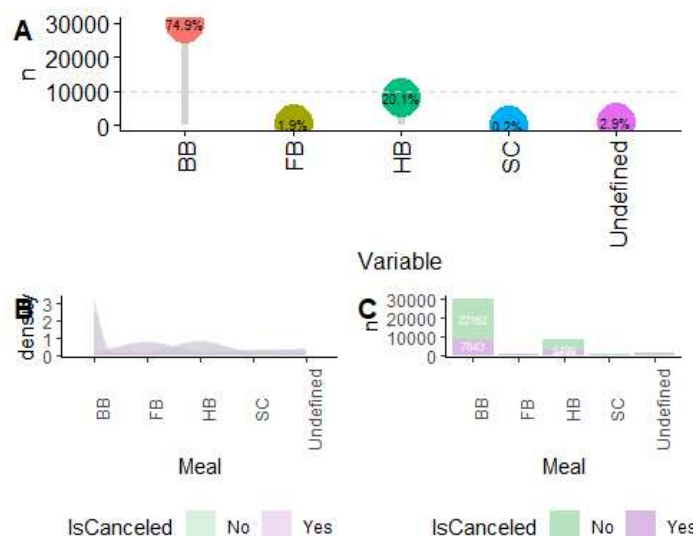
Con respecto a la variable “**Required Car Parking Spaces**”, se muestra que los que solicitan algún espacio son más probables de no cancelar. (Ver Fig.5).

**Fig: 5.**



La mayoría de reservas han elegido la opción “BB” para la variable “**Meal**”. La proporción entre canceladas y no canceladas es casi equitativa (Ver Fig.6).

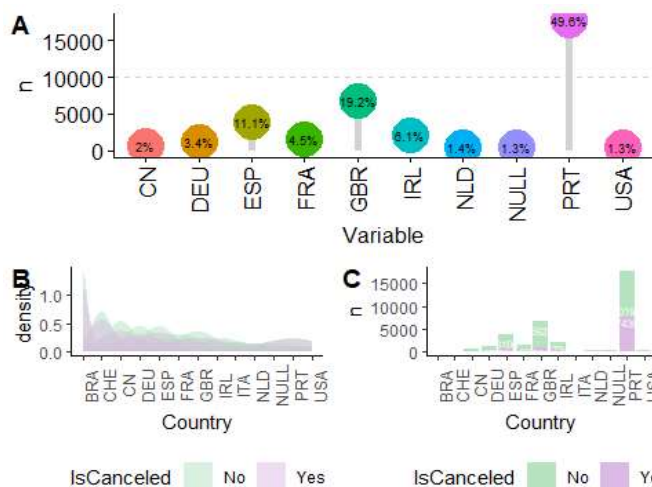
**Fig: 6.**



El **país** con mayor número de reservas es Portugal

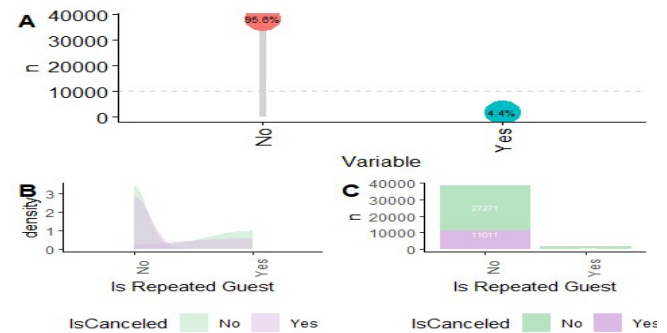
y también se observó que es el que tiene alto número de cancelaciones. Algarve no es un destino turístico tan popular como otras ciudades (ciudades cercanas al mar Mediterráneo), y la preferencia por esta ciudad es mayor en los residentes de Portugal. Sin embargo, consideramos que esta información es subjetiva porque, al momento de realizar la reserva, algunos usuarios prefieren no indicar su verdadera residencia o por error, y el hotel tiene la información correcta en el checkin. Por otro lado, la diferencia entre canceladas y no canceladas no es significativa (Ver Fig.7).

Fig: 7.



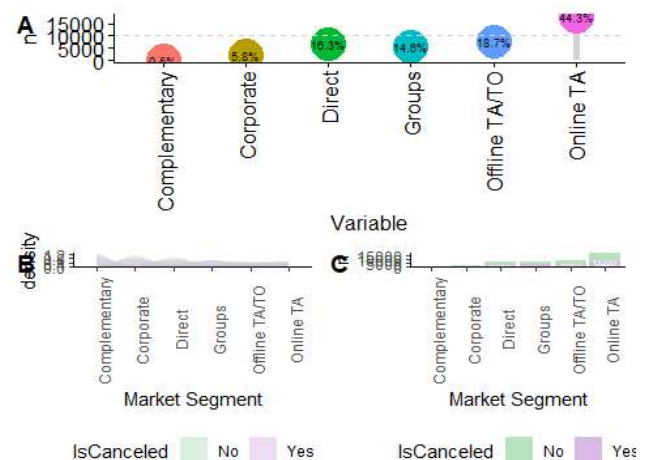
La mayoría de huéspedes son clientes nuevos, y cabe resaltar que tienen más incidencia de no cancelar. Esto puede deberse a las campañas de publicidad (con opción más económica de alojamiento) y a los OTAs (Booking, Expedia entre otros). Al ser económicos no tienen la opción de cancelación o devolución (estrategias rígidas para reducir las cancelaciones) (Ver Fig.8).

Fig: 8.



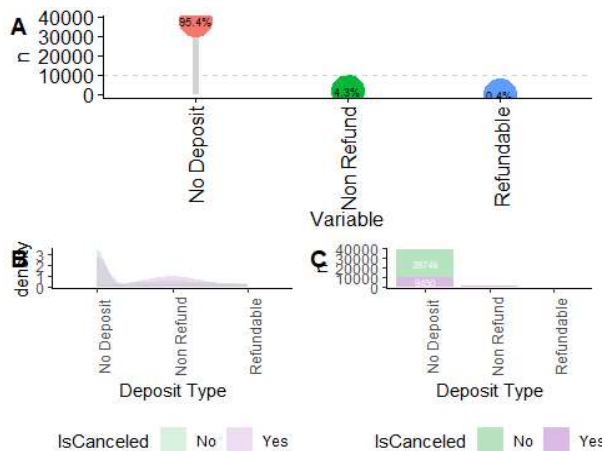
Una gran proporción de reservas se realizó mediante un OTA. Los estudios nos habían anticipado que la tendencia actualmente es reservar mediante algún OTA por sus técnicas de marketing (Ver Fig.9).

Fig: 9.



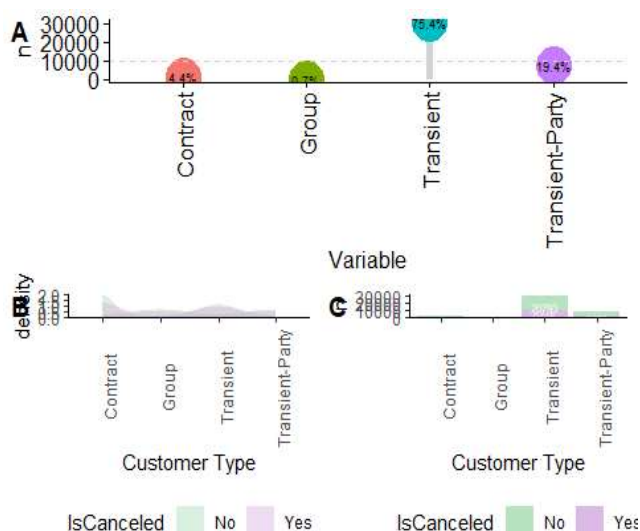
Existe una gran proporción de Reservas tipo "No Deposit", puede deberse a que la mayoría de usuarios que reserva por OTA elige las opciones de "Pagar después" o "Pagar en establecimiento" (Ver Fig. 10)

**Fig: 10.**



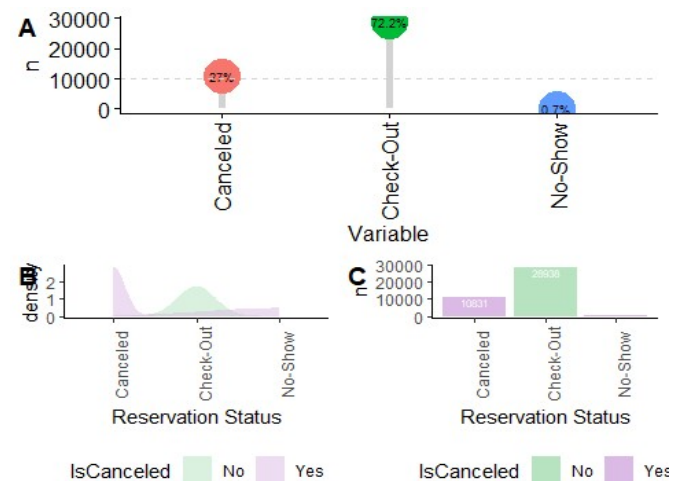
Hay una significativa proporción de clientes tipo "Transient" (75.4%). (Ver Fig. 11)

**Fig: 11.**



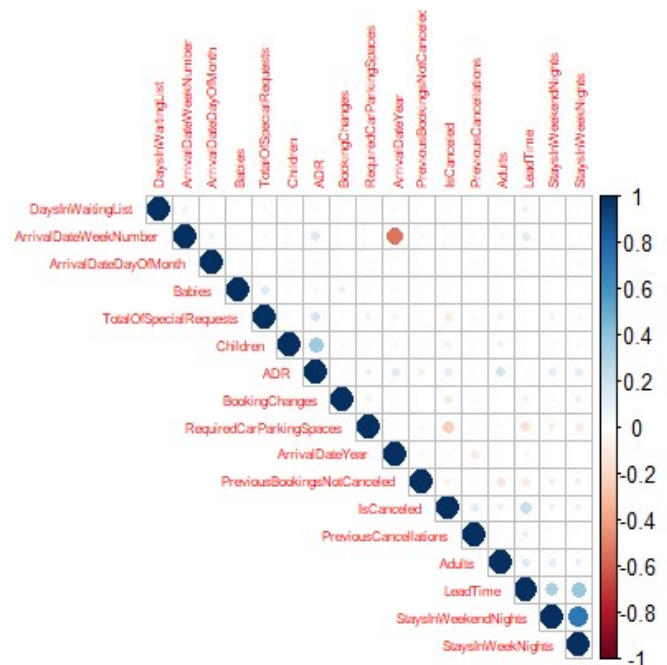
Si comparamos la variable "Reservation Status" con "Is Canceled" tenemos observaciones razonables, por ejemplo, todas las reservas que resultaron checkout son No canceladas, en cambio, las canceladas contiene las que su último estado fue Canceled o No Show (Ver Fig. 12)

**Fig: 12.**



Para finalizar encontramos que las variables **StaysInWeekendNights** y **StaysInWeekNights** están altamente correlacionadas positivamente, lo cual refleja que un significativo número de reservas se realizan sin diferencia si es entre semana o fines de semana.

**Fig: 13.**



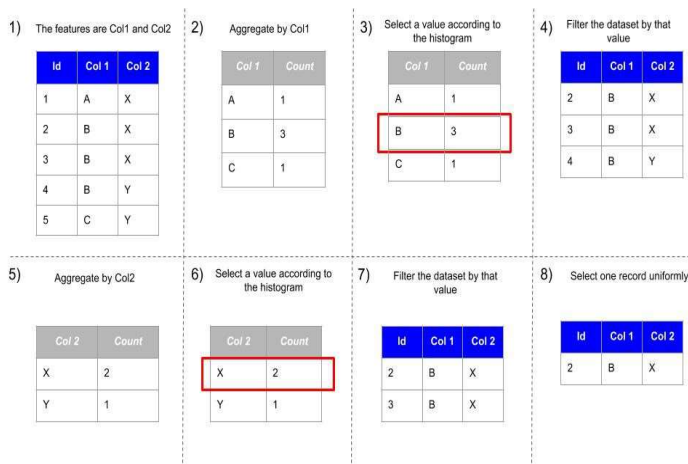
Después se realizó la preparación de los datos para la construcción de los modelos. Se extrajo las variables **ReservationStatus** y **ReservationStatusDate** por que guardan información actualizada que no se tendrá definida al momento de que ingresar la reserva.

La variable **ArrivalDateMonth** es transformada en variable ordinal en base al orden de los meses.

En el caso de las variables del tipo carácter (**CustomerType**, **DepositType**, **DistributionChannel**, **MarketSegment**, **Meal**), se convirtieron en categóricas y se agregaron las columnas dummy (de carácter dicotómico:0-1).

Las variables **AssignedRoomType**, **Country** y **ReservedRoomType** tienen más de 5 niveles, por lo que convertir en columnas dicotómicas y agregarlas al dataset es ineficiente (por ejemplo, en el caso de **Country** tiene 125 niveles de factor lo que implica agregar 124 columnas, no es óptimo). Por ello, se utiliza la función WOE para convertir los valores categóricos a numéricos en base a su peso con respecto a la variable objetivo (**IsCanceled**) [4]. (Ver Fig. 14)

**Fig: 14.**



Las variables **Agent** y **Company** guardan los códigos de los agentes y compañías de viajes, pero tienen una categoría NULL, para reducir la complejidad de los datos se agregó un código 999 que identifica estos casos.

Finalmente se almacenan los datos codificados en un archivo csv, que son utilizados en las técnicas de agrupación y predicción.

### OPERACIONES ESPECIFICAS A LA ASIGNATURA TECNICAS DE AGRUPACION Y REDUCCION DE LA DIMENSION

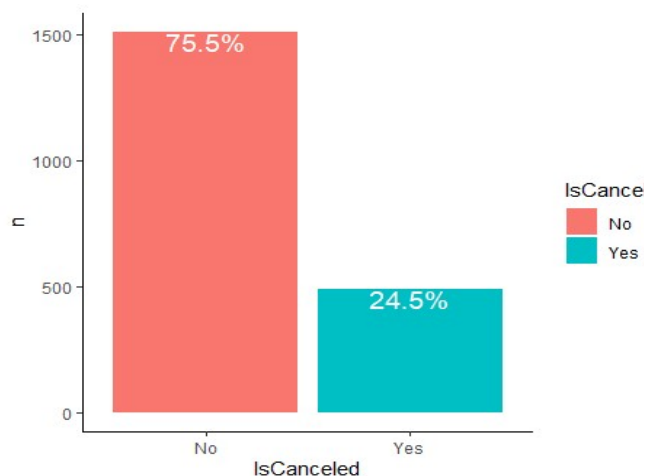
Para la asignatura de Técnicas de Agrupación y Reducción de la Dimensión, se necesitan procesar algoritmos que requieren una gran capacidad computación (Matrices de Distancias y Clustering), es por ello, que se generaron dos muestras:

- De los datos codificados (todas las variables del tipo numérico).
- De la data original, esta utilizada para las matrices de correspondencia.

Con la finalidad de que la muestra explique significativamente la varianza de los datos, se utilizó la técnica estratificada en base a ponderaciones de los histogramas con respecto a la variable objetivo (**IsCanceled**), propuesta por Gianluca Malato en Towards Data Science[4]. Muestra Codificada (Ver Fig.15)

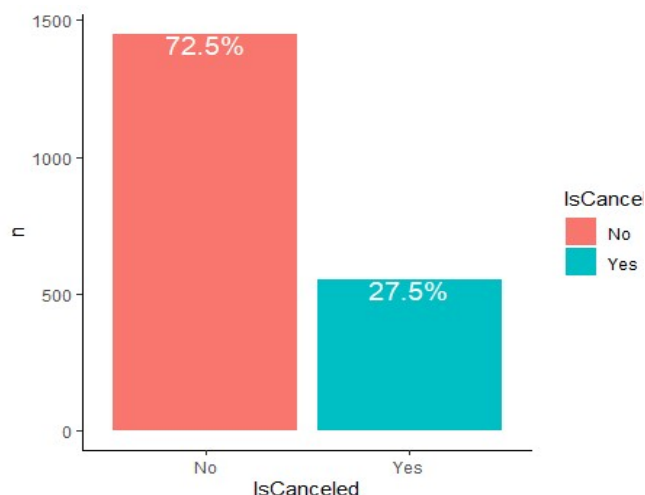
**Fig: 15.**





Muestra No Codificada (Variables numéricas y categóricas para el análisis de correspondencias) (Ver Fig.16)

**Fig: 16.**



Finalmente se guardaron las dos muestras en archivos csv.

### OPERACIONES ESPECIFICAS A LA ASIGNATURA PREDICCION

En cuanto al modelo predictivo de regresión para predecir cuán probable es que “la reserva se cancele”, utilizaron los datos codificados y guardados en las operaciones comunes.

En cuanto al modelo predictivo para el pronóstico del número semanal de las reservas, se necesita previamente preparar los datos y que se construya la siguiente estructura:

Fecha de Reserva	Número de Reservas
05/11/2014	12

Para el cálculo de estos datos se consideró lo siguiente:

- Fecha de Reserva: es la fecha resultante de la Fecha de arribo menos los días de Lead Time.
- Número de Reservas: se cuantifica las observaciones (reservas) realizadas en el día de reserva.

La información de las reservas demuestra que es una serie temporal irregular (que los registros tienen intervalos de tiempo sin registros). Dado que los modelos de ARIMA buscan las relaciones de las observaciones con respecto a sus valores pasados, es necesario imputar la serie temporal (convertirla en regular). Se utilizó la función complete de la librería dplyr y se imputaron las filas con valor 0 (considerando que no se realizaron reservas esas semanas)[6,7]. Finalmente, se almacenó los datos en un archivo csv para su procesamiento en el informe 3 de Predicción.

### CONCLUSIONES

La información proporcionada guarda mucha concordancia con las situaciones reales de los hoteles, se demostraron en el análisis exploratorio de los datos.

En cuanto al trabajo de preparación de datos, como el conjunto de datos es extenso (40060 registros con 31 variables) se debieron hacer codificaciones a las variables categóricas para que el procesamiento del desarrollo del prototipo no se extendiera del tiempo límite del proyecto [5].

En los sitios de internet, se encontraron muchas técnicas distintas de codificación y limpieza sin

embargo los métodos “WOE” y “Dummyfication” son los más recomendados para el caso del proyecto.

La técnica de muestreo basado en ponderaciones calculo una muestra muy adecuada para las técnicas de agrupación y segmentación.

### REFERENCIAS BIBLIOGRAFICAS

- [1] Study: Cancellation rate at 40% as OTAs push free change policy, HotelManagement.net ([link](#))
- [2] Los hoteles ante el aumento de las cancelaciones, ¿Qué hacer?, Tecnohotel, ([link](#))
- [3] A data scientist toolkit to encode categorical variables to numeric, ([link](#))
- [4] Stratified sampling and how to perform it in R, ([link](#))
- [5] Clustering Mixed Data Types in R, ([link](#))
- [6] Time Series Modelling with irregular data, ([link](#))
- [7] Everything you need to know about time series ([link](#))

### ANEXO

- Archivo RMarkdown en [Github](#) (MasterDS2019>Prediccion>Research>Examen20-Informe1MGN.Rmd)