

Informe del Análisis Exploratorio de Datos para la Cancelación de Reservaciones de Hoteles

**Preparado por
Luis Rubén Artavia**

Descripción General

Para un hotel es muy importante lograr tener una ocupación del 100% o al menos un porcentaje bastante cercano a este, pero como parte de este proceso tiene que lidiar con la posibilidad de que las reservaciones que son hechas pueden ser canceladas y dejar de percibir el ingreso respectivo.

En algunos casos, las reservaciones pueden requerir de un depósito o asegurarse un monto en caso de una cancelación fuera del tiempo de gracia otorgado por el hotel, pero a pesar de esto, si ese espacio logra ser utilizado, la ganancia es mayor.

El objeto del análisis de este trabajo es poder identificar patrones que podrían ser de ayuda a los encargados de los hoteles y así definir políticas que permitan maximizar el uso de las habitaciones.

Es importante recalcar que este estudio no considerará eventos especiales, como el caso de la pandemia del Covid-19 que se está viviendo en este momento.

Exploración de los Datos

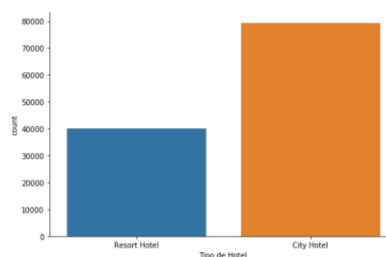
La información que fue utilizada para trabajar en la resolución del problema incluye los siguientes datos:

- Los hoteles se clasificaron en Urbanos (City) y Turísticos (Resort)
- Los datos de la demanda de reservaciones de hoteles corresponde a los años 2015, 2016 y 2017
- Incluye información de cuando realizó la reservación, la duración de la estadía, cantidad personas divididas en adultos, niños y bebés
- El tiempo transcurrido desde la fecha de reservación hasta la fecha de arribo, tipo de alimentación
- El tipo de depósito realizado, canal de distribución, días en espera, tipo de cliente
- El indicador para identificar si la reservación fue cancelada o no

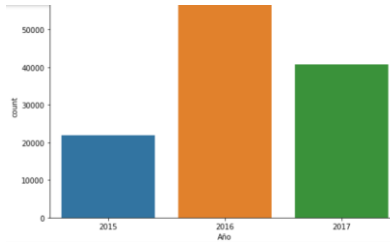
De la exploración, se considera que los atributos que pueden ser considerados como significativos para el problema que se está tratando, tienen que ver con el tipo de hotel, el tipo de comida reservada, cantidad de días de hecha la reservación, mes de arribo de la reservación, segmento del mercado, canal de distribución, tipo de depósito y tipo de cliente. Por otro lado, según la matriz de correlación, los atributos de tiempo de hecha la reservación, estadía, cantidad de adultos, si hay cancelaciones anteriores, días en lista de espera y costo promedio de la reservación, podrían ser usados en el proceso de para definir el modelo predictivo.

De la exploración de datos, tenemos los siguientes resultados:

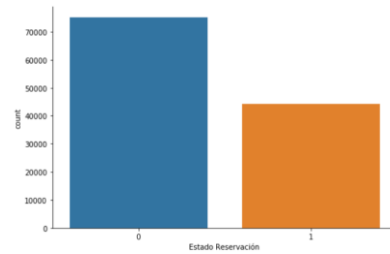
La reservación en hoteles urbanos representa un 66% del total y los hoteles turísticos un 34%:



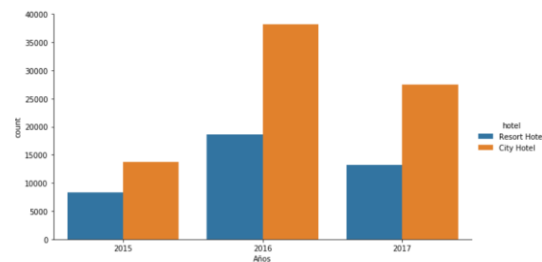
La información recolectada corresponde a un 18% de reservaciones en el año 2015, 47% en el 2016 y 34% en el 2017:



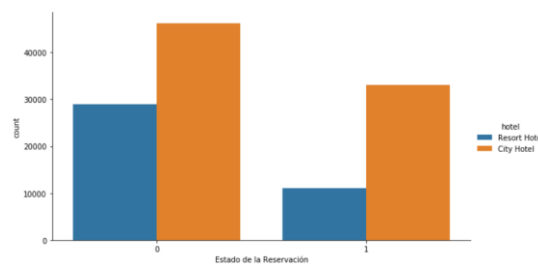
Con relación a la definición de las cancelaciones, tenemos en general un 37% sobre las reservaciones que fueron recolectadas en los datos utilizados:



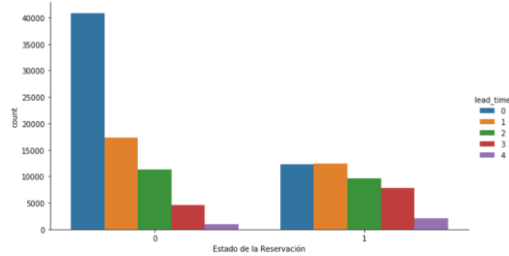
Se observa un leve aumento en la predilección de los hoteles urbanos sobre los turísticos. Para el año 2015 la relación es 62% urbanos y 38% turísticos, en el 2016 67% urbanos y 33% turísticos, y en el 2017 68% urbanos y 32% turísticos:



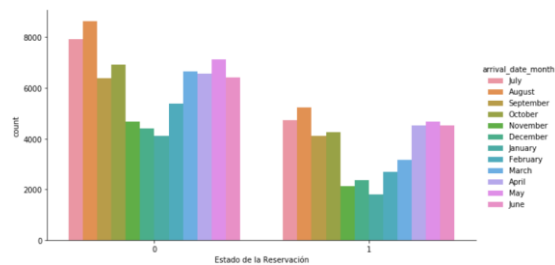
A pesar del leve aumento en los hoteles urbanos, el porcentaje de cancelaciones en este tipo de hotel es mayor al turístico. Tenemos un 42% de cancelaciones en los urbanos y apenas un 28% en los turísticos, por debajo del porcentaje general:



El tiempo transcurrido entre la fecha de reservación y la fecha de arribo es un factor importante en la valoración de las cancelaciones. Entre menos días se tienen, menor es el porcentaje de cancelaciones. Tenemos que cuando son menos de 147 días es un 23% que además está por debajo del porcentaje general, cuando son más de 147 pero menos de 295 sube a un 42% las cancelaciones, cuando es superior a 295 pero menor a 443 días el porcentaje es 46%, entre 444 y 591 tenemos un 63% y más de 591 días sube a 68%:



En cuanto al mes de arribo al hotel, tenemos que los meses de enero, febrero, marzo, julio, noviembre y diciembre son iguales o están por debajo del valor general. En resumen, los porcentajes son: Enero 30%, Febrero 33%, Marzo 32%, Abril 41%, Mayo 40%, Junio 41%, Julio 37%, Agosto 38%, Setiembre 39%, Octubre 38%, Noviembre 31% y Diciembre 35%:



Métodos de Clasificación

Un método que puede ser utilizado para descubrir más información, es la aplicación de algoritmos de clasificación de información. Estos procesos complementarían el análisis manual que realizan los Científicos de Datos de una manera automatizada. En nuestro caso, se probaron los métodos de Support Vector Classification, Random Forest Classifier y Ridge Classifier, obteniendo el que mejor resultados en la etapa de entrenamiento el Random Forest Classifier.

Observaciones

Al realizar comparaciones de las diferentes características de las reservaciones con respecto a las cancelaciones, tenemos las siguientes conclusiones:

- En los hoteles de turismo la cantidad de cancelaciones son menores a los urbanos, 28% y 42% respectivamente, pero vemos que la cantidad de reservaciones en los hoteles urbanos pasó de un 62% a un 68% en dos años.
- El tipo de comida asociada a la reservación no influye en las cancelaciones, incluso si no las incluye, pues están entre el 34% y el 37%. El único factor que podría impactar negativamente es cuando se incluyen todas las comidas, porque las cancelaciones llegan al 60%.
- Para el caso de la cantidad de días transcurridos desde el momento que se ingresó la reservación hasta la fecha de arribo, cuanto más cerca están menor es el porcentaje de cancelaciones. Cuando sobre pasan los 148 días los valores sobrepasan el porcentaje general, van desde el 42% hasta el 68%.

- Teniendo como base el porcentaje general de cancelaciones, se podría decir que los meses donde el porcentaje de cancelación está por debajo son aquellos donde hay más movimiento de turistas o están las vacaciones escolares de muchos de los países.
- Con respecto al canal de distribución, el porcentaje más bajo es cuando el cliente hace la reservación directa, 17%, y con intermediario es el más alto, 42%.

Se podría considerar que las opciones de comida no vienen a ser un factor importante para el cliente a la hora de mantener la reservación activa , que las reservaciones hechas con poca anticipación son más seguras y que tener la opción de que el cliente pueda hacer la reservación directamente mejora la ocupación del hotel.

Por otro lado, cuando se presentan los factores negativos, el hotel podría considerar el aceptar un porcentaje sobre la ocupación máxima y así tener una alta ocupación del lugar.