



Instituto Tecnológico
de Buenos Aires

82.05 - Análisis Predictivo

CASO DE ESTUDIO

PREDICCIÓN DE CANCELACIONES DE RESERVAS HOTELERAS

-

SOL ALEJANDRA WINKEL

CONTENIDOS

01 Introducción

Establecer objetivos del trabajo.

03 Análisis Exploratorio

Presentar gráficamente los datos para conocer mejor las variables.

02 Preparación Base

Seleccionar la base y preparar las variables.

04 Conclusiones

Utilizar lo realizado previamente para obtener conclusiones respecto a la base.





OBJETIVO

El objetivo principal es entrenar un modelo predictivo que permita a los hoteles utilizar la totalidad de las habitaciones considerando incluso las reservas que podrían ser canceladas.

Poder predecir quién cancelará la reserva a partir de distintas características de la reserva. como: cantidad de adultos, cantidad de niños, cantidad de noches, tipo de habitación, fecha, etc.

Un modelo de respuesta puede proporcionar un aumento significativo en el rendimiento del hotel al tener la ocupación al máximo y aumentar las ganancias.

BASE DE DATOS

La base Hotel Reservation fue obtenida de Kaggle. Tiene 19 variables (15 numéricas y 4 de tipo carácter). Contiene 36.275 registros de reservas.

Variables

- Booking_ID: identificador único de cada reserva
- no_of_adults: cantidad de adultos
- no_of_children: cantidad de niños
- no_of_weekend_nights: cantidad de noches de fin de semana (Sábado o Domingo) que el cliente reservó para quedarse.
- no_of_week_nights: cantidad de noches de la semana (Lunes a Viernes) que el cliente reservó para quedarse.
- type_of_meal_plan: tipo de menú que el cliente reservó
- required_car_parking_space: El cliente necesita una cochera? (0 en caso de que No, 1 en caso de que Si)
- room_type_reserved: Tipo de habitación reservada por el cliente. Los valores son codificados por INN Hotels.
- lead_time: Número de días entre la fecha de reserva y la fecha de llegada
- arrival_year: año de la fecha de llegada

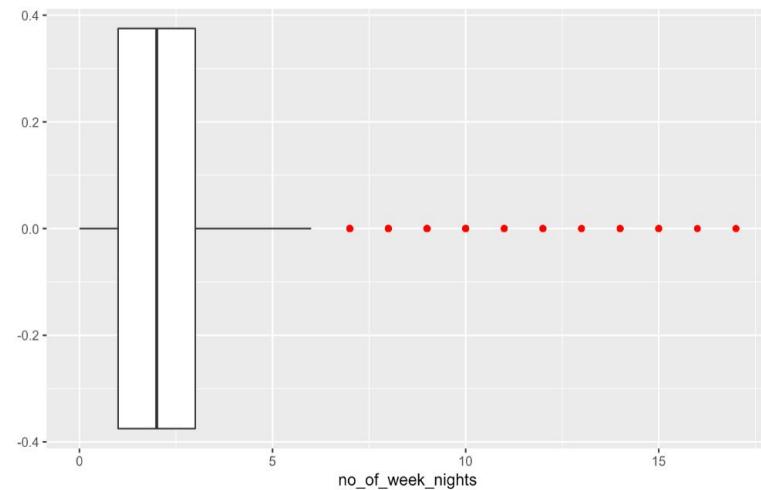
- arrival_month: mes de la fecha de llegada
- arrival_date: día del mes de la fecha de llegada
- market_segment_type: designación del segmento de mercado.
- repeated_guest: es el cliente un huésped repetido? (0 en caso de que No, 1 en caso de que Si))
- no_of_previous_cancellations: Número de reservas anteriores canceladas por el cliente antes de la reserva actual
- no_of_previous_bookings_not_canceled: Número de reservas anteriores no canceladas por el cliente antes de la reserva actual
- avg_price_per_room: Precio medio por día de la reserva (en euros)
- no_of_special_requests: Número total de solicitudes especiales realizadas por el cliente (por ejemplo, piso alto, vista desde la habitación, etc.)
- Booking_Status: variable target que indica si la reserva fue cancelada o no

CAMBIOS EN EL DATASET

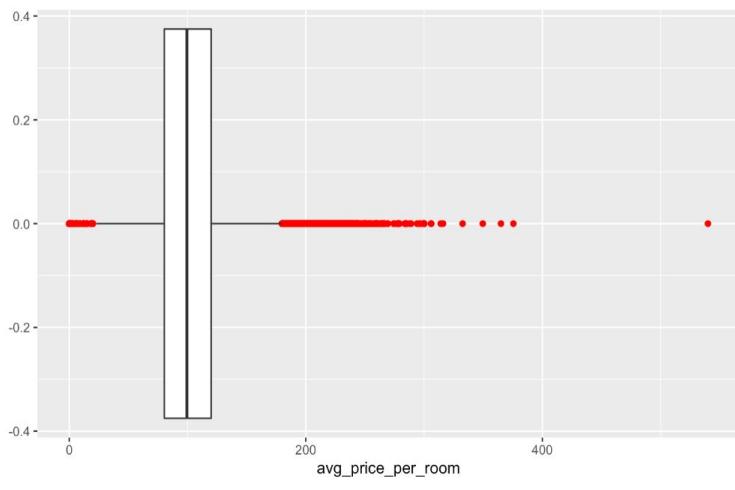
01	Verificar el formato	<ul style="list-style-type: none">• ID como factor, y no como número• Día, mes y año como enteros
02	Creación de nuevas variables	<ul style="list-style-type: none">• Total de huéspedes: adultos + chicos• Fecha de llegada: día/mes/año
03	Análisis de datos faltantes	<ul style="list-style-type: none">• Las variables no poseen NAs

Booking_ID	0	no_of_adults	0	no_of_children	0
no_of_weekend_nights	0	no_of_week_nights	0	type_of_meal_plan	0
required_car_parking_space	0	room_type_reserved	0	lead_time	0
arrival_year	0	arrival_month	0	arrival_date	0
market_segment_type	0	repeated_guest	0	no_of_previous_cancellations	0
no_of_previous_bookings_not_canceled	0	avg_price_per_room	0	no_of_special_requests	0
booking_status	0				

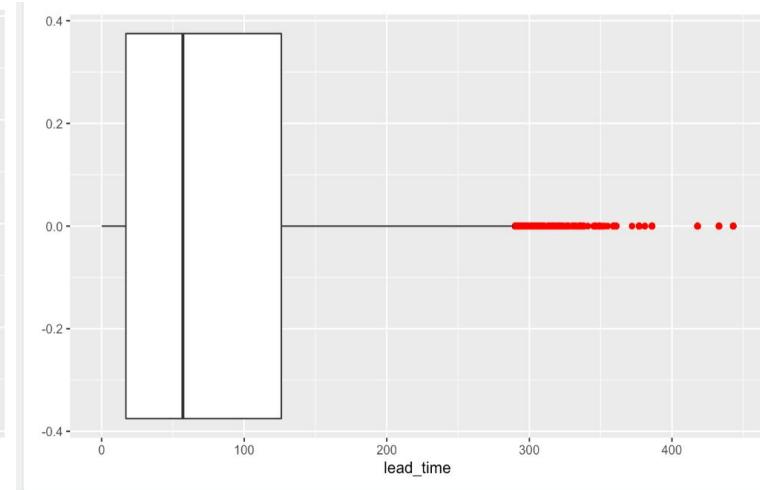
OUTLIERS (valores atípicos)



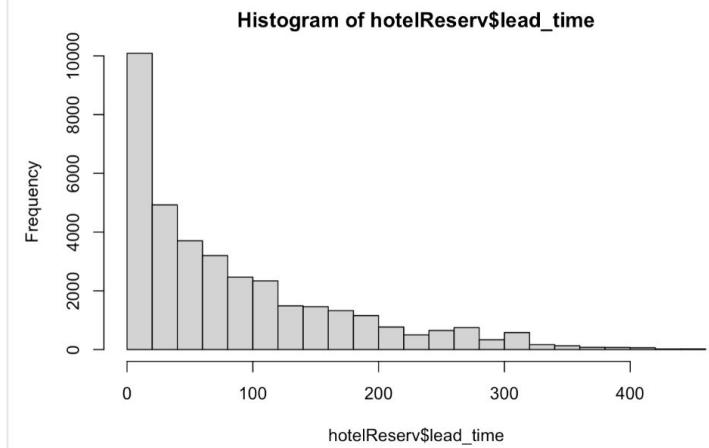
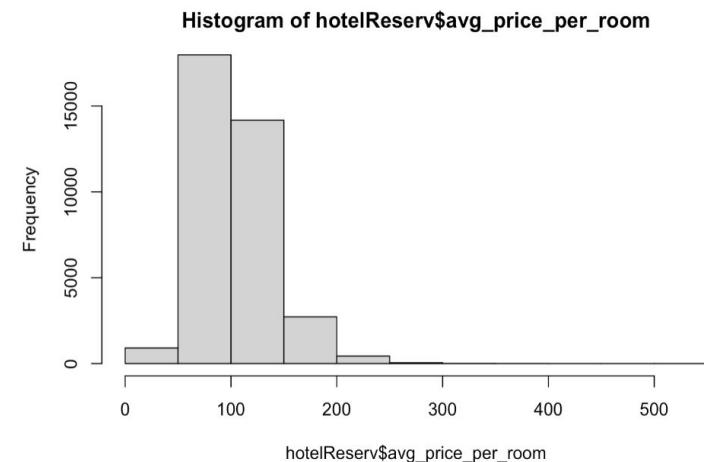
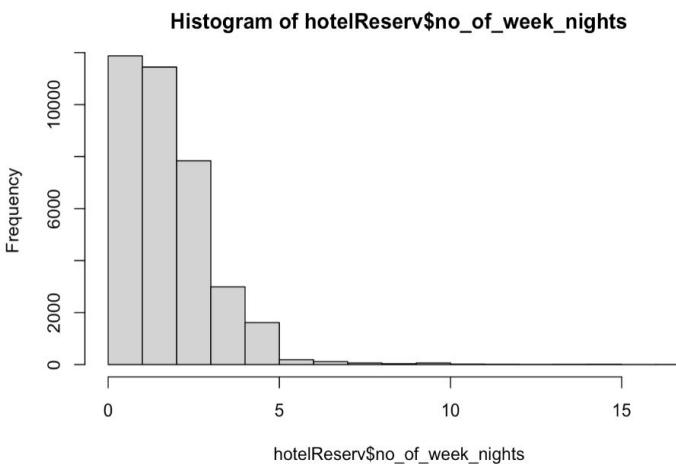
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
<code>no_of_week_nights</code>	0.000	1.000	2.000	2.204	3.000	17.000



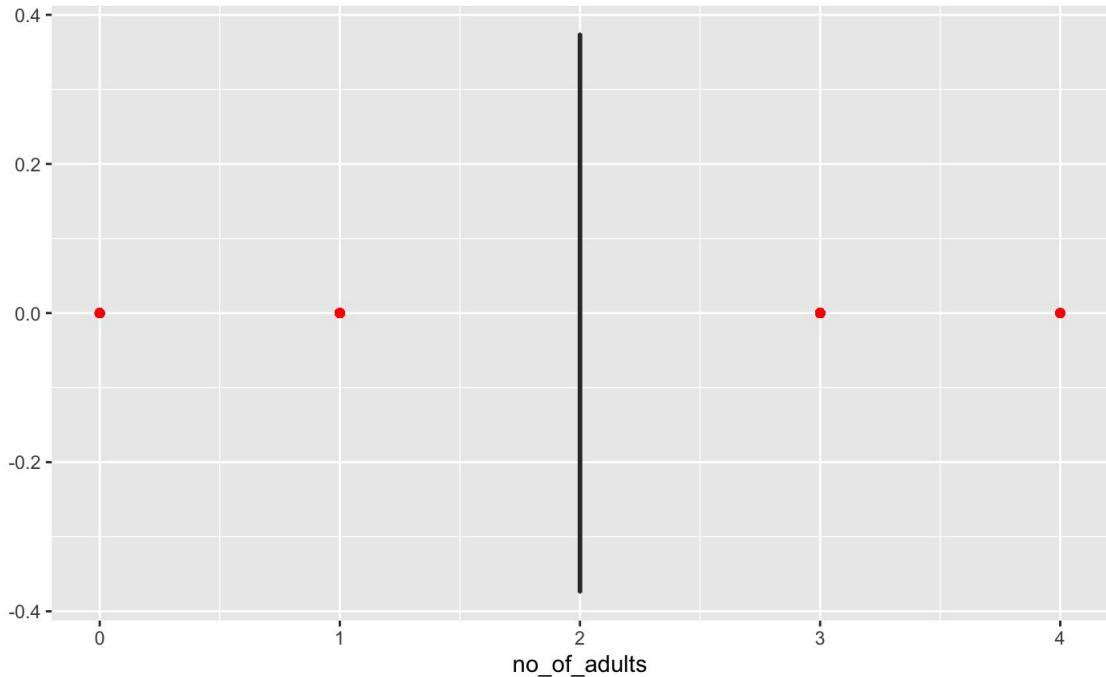
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
<code>avg_price_per_room</code>	0.00	80.30	99.45	103.42	120.00	540.00



	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
<code>lead_time</code>	0.00	17.00	57.00	85.23	126.00	443.00



OUTLIERS (valores atípicos)

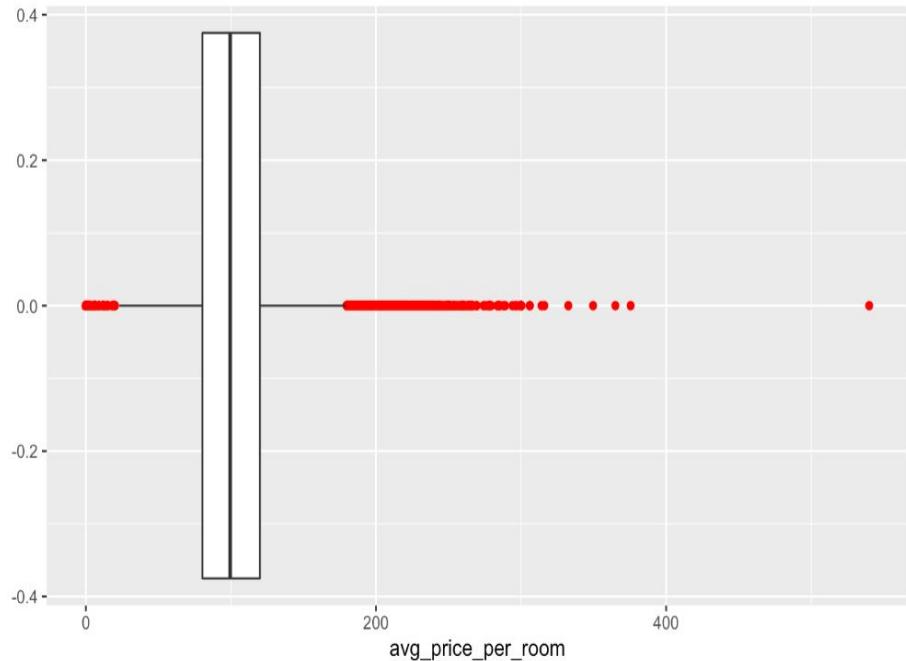


Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	2.000	2.000	1.845	2.000	4.000

Los valores se encuentran entre 0 y 4.

- Hay 139 registros que no tienen adultos en sus reservas pero sí chicos

OUTLIERS (valores atípicos)

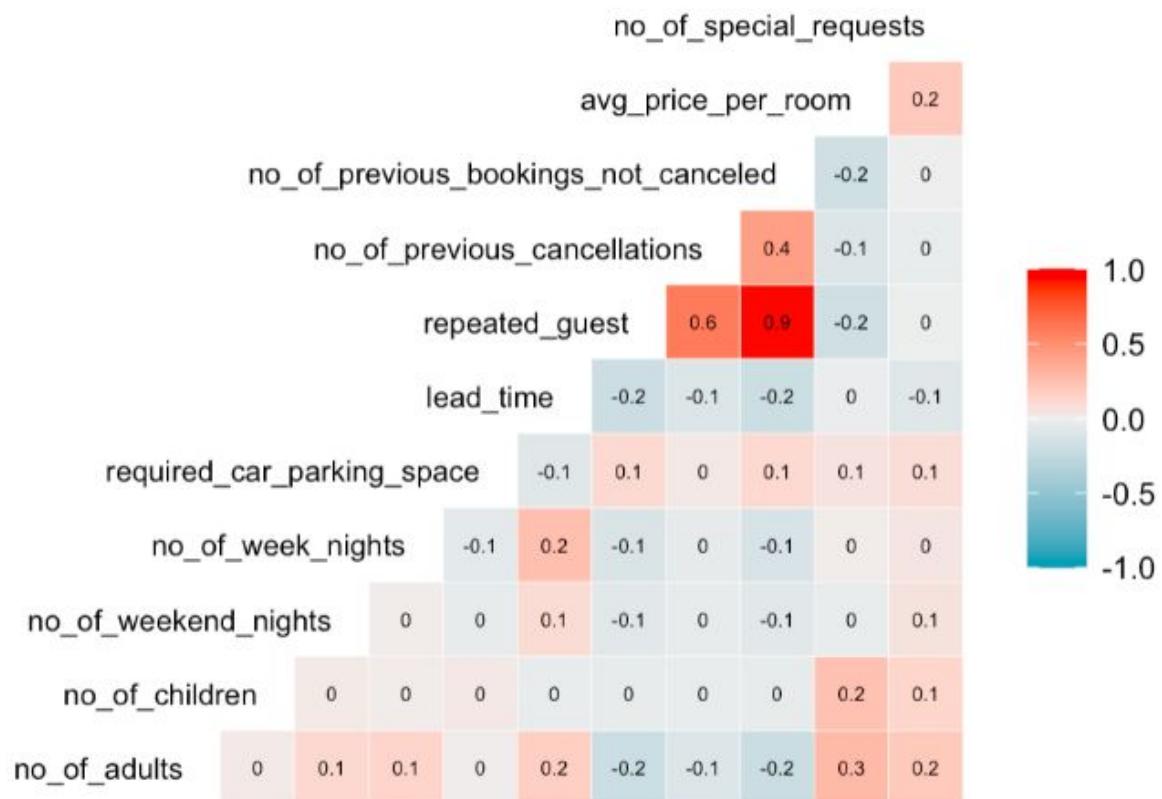


Cantidad de registros `avg_price_per_room = 0` → 545

- Complementary → 354
- Online → 191

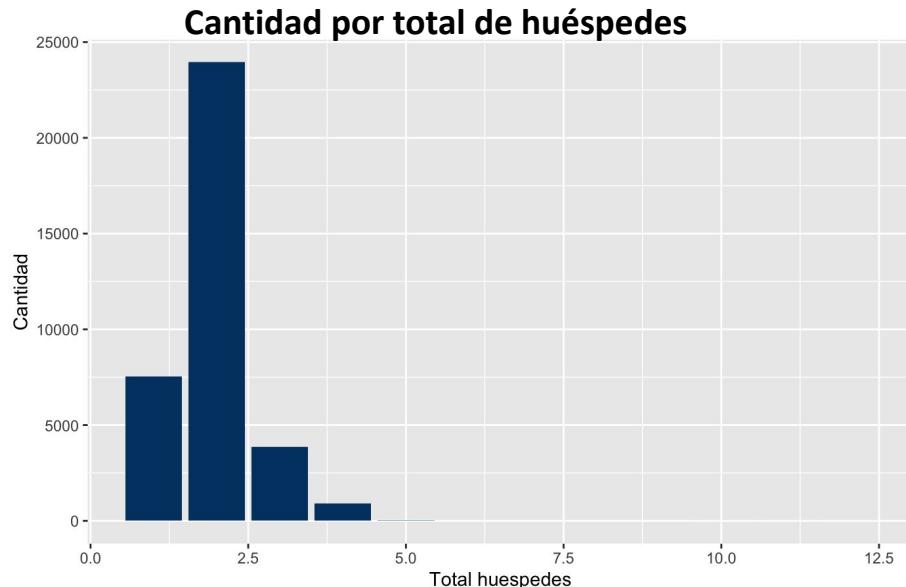
Se imputan por la media los registros de “Online”

ANÁLISIS DE CORRELACIÓN



Las variables correlacionadas positivamente están en rojo y las variables negativamente correlacionadas están en azul.

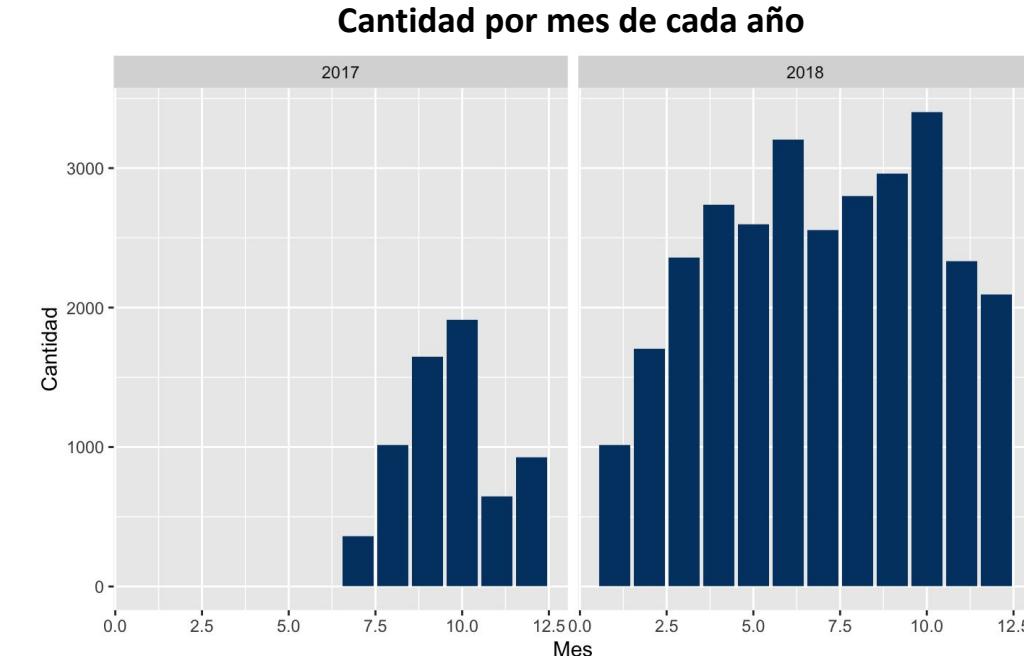
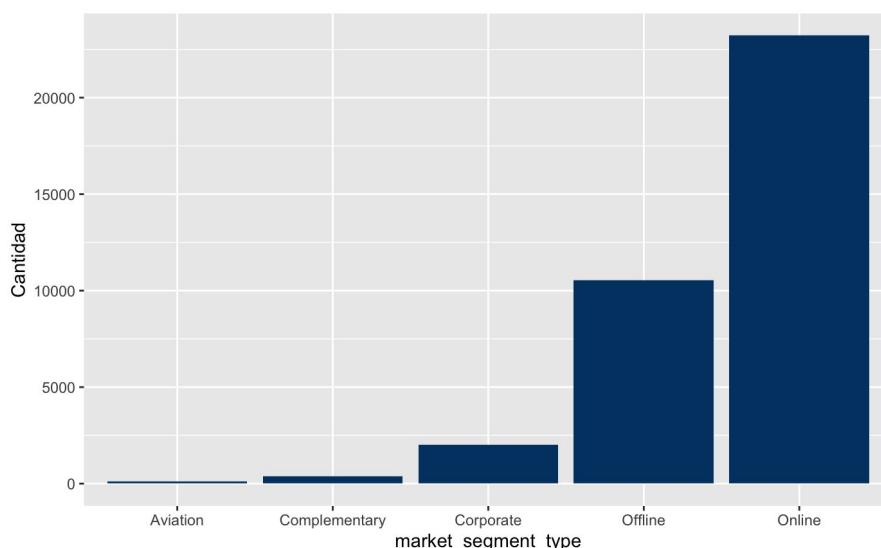
Análisis de las reservas



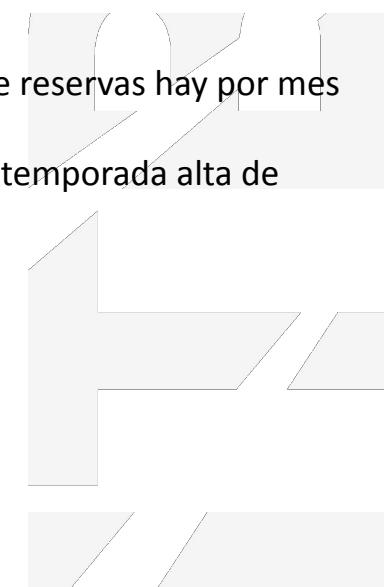
Que cantidad de reservas hay por segmento de mercado
De donde provienen la mayor cantidad de las reservas

Que cantidad de huéspedes se alojan por reserva

Cantidad por segmento del mercado

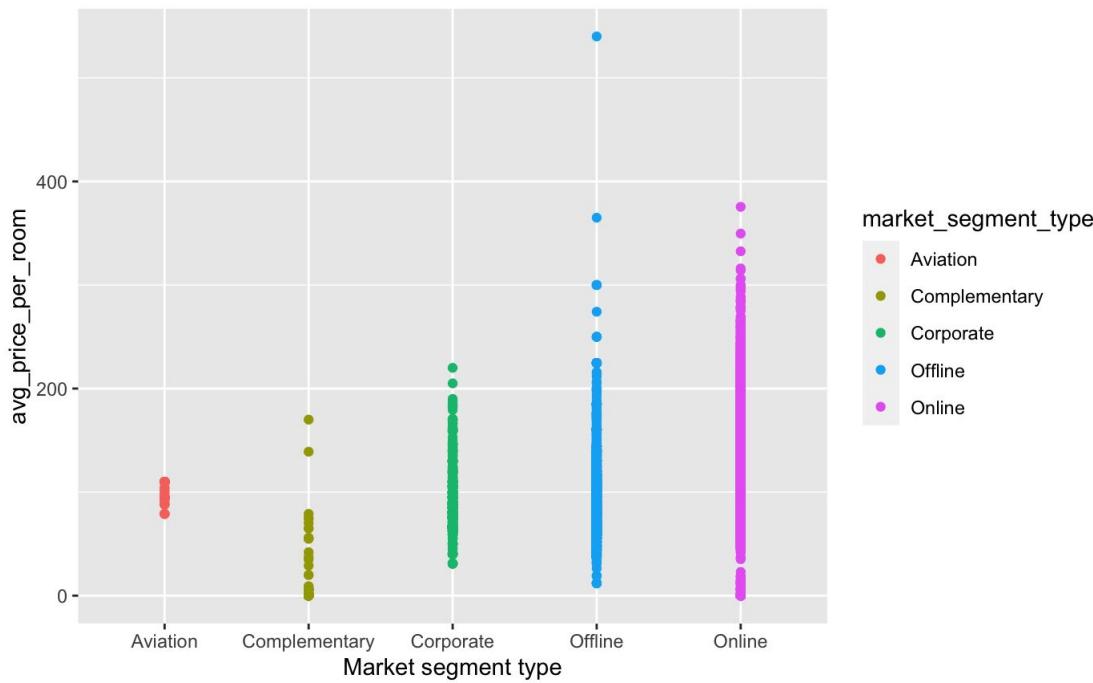


Que cantidad de reservas hay por mes de cada año
Qué meses son temporada alta de reservas



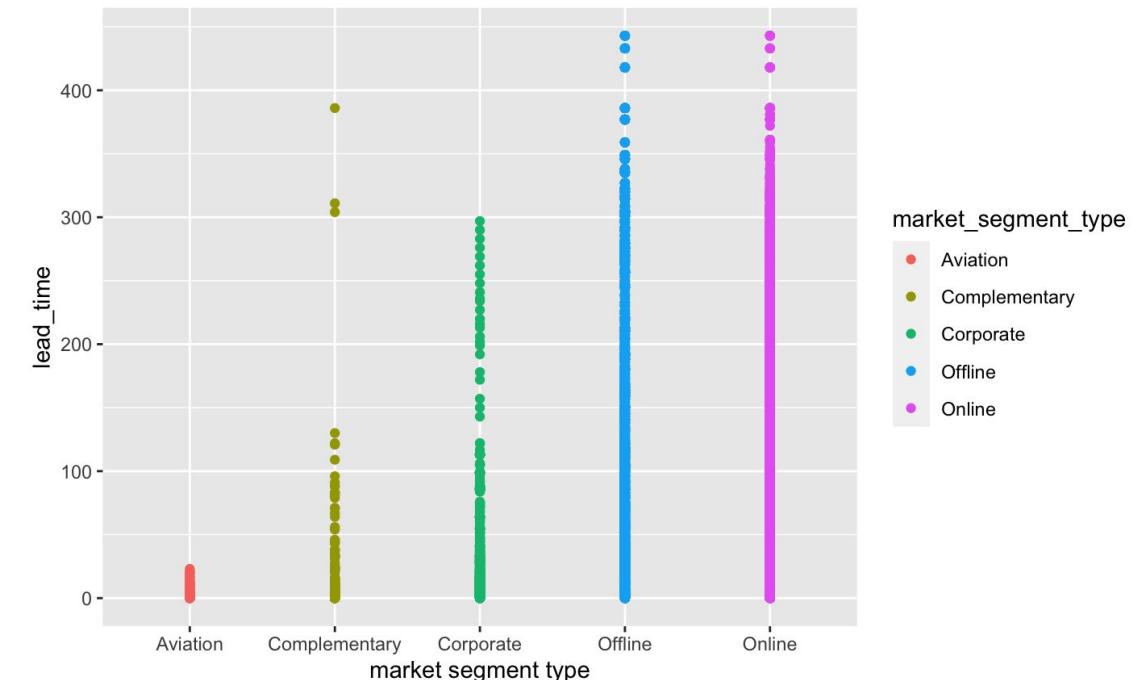
Análisis de los segmentos de mercado

Precio promedio por noche por segmento de mercado



Existe una relación entre el segmento y el precio de la noche que reserva?

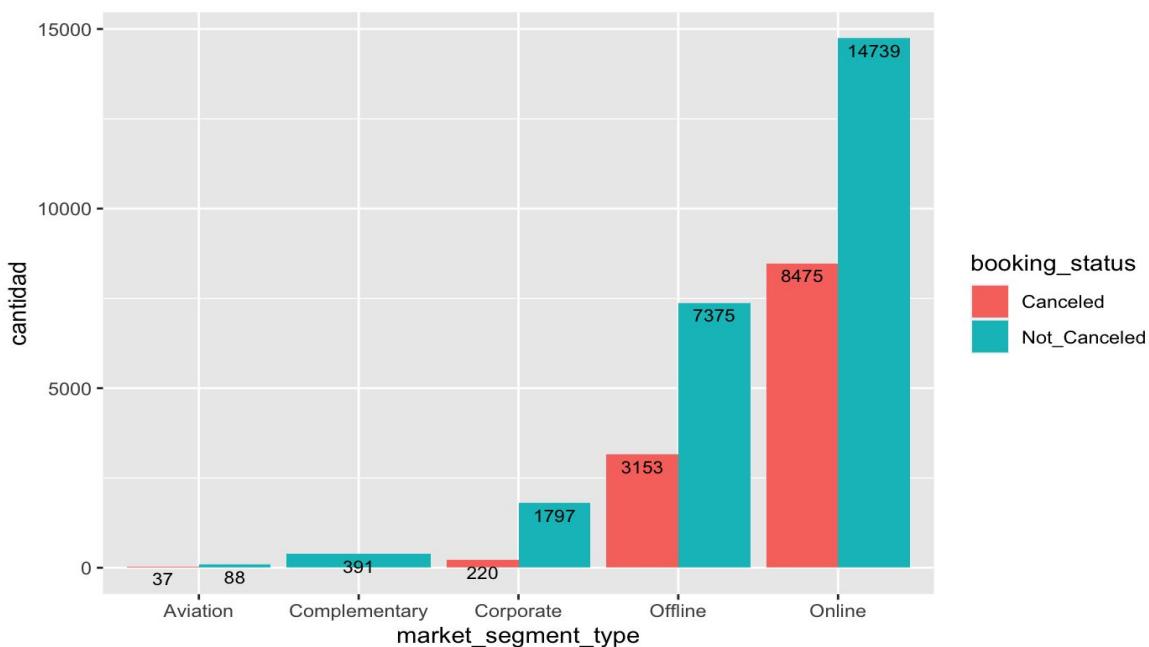
Días de anticipación en la realización de la reserva por segmento de mercado



Cómo se relacionan los días de anticipación de la reserva con el segmento del mercado

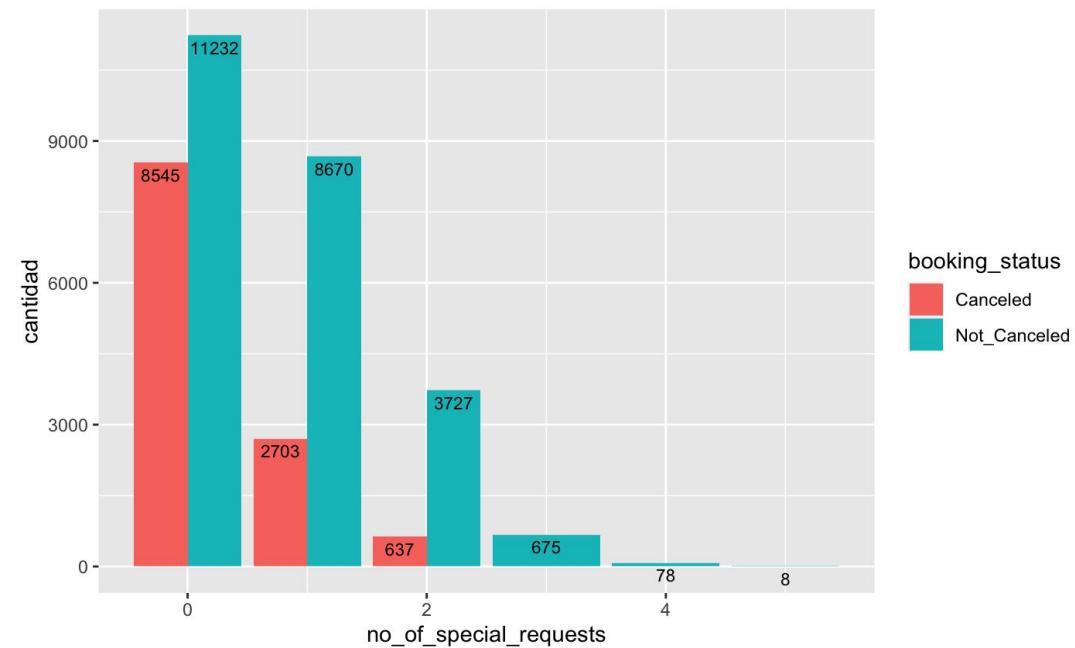
Cancelación de las reservas

Cantidad de reservas por segmento



Cuántas reservas fueron o no canceladas, dependiendo del segmento del mercado.

Cantidad de reservas por cantidad de requerimientos especiales



Visualizar por la cantidad de requerimientos especiales, si la reserva fue cancelada o no

CONCLUSIONES

- El dataset tiene un gran potencial para su continuo uso, considerando la facilidad que tiene para ir actualizándose así como también para incorporar nuevas variables que resulten relevantes.
- Es una base que puede ser aplicada por empresas hoteleras
- A futuro se podrían agregar nuevas variables que sean necesarias para poder indicar en caso de cancelación, si volverá a reservar o no en esa empresa.
- En cuanto a lo observado, se podría decir que no hay una fuerte correlación entre las variables.
- Así como también se pudo ver que las reservas de personas que se alojan por trabajo son menos propensas a cancelar las reservas, junto con las reservas que tienen al menos 2 requerimientos especiales.



Instituto Tecnológico
de Buenos Aires

Muchas gracias