

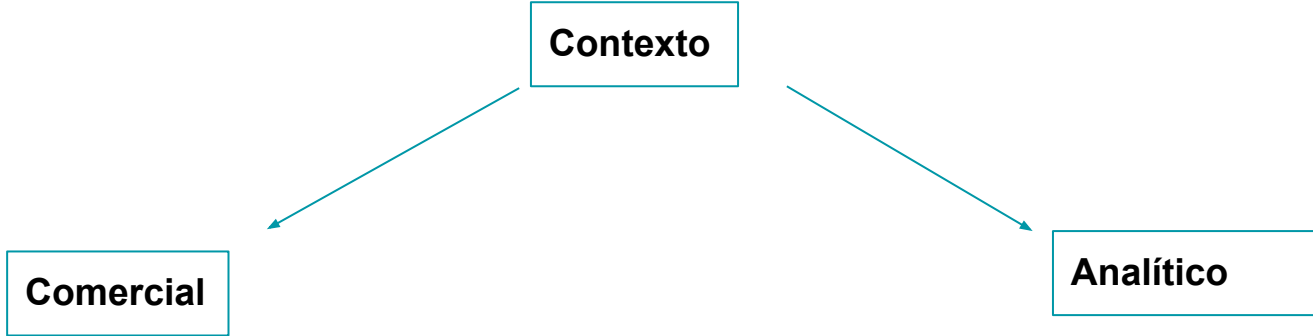
# **Análisis de canales de reserva de hoteles**

(estudio sobre la evolución de la reserva de hoteles y sus cancelaciones para los años 2017-2018)

- 1. Contexto**
- 2. Objetivos y preguntas de Interés**
- 3. Análisis Exploratorio de datos**
- 4. Obtención de Insights**



Los canales de reserva de hoteles han experimentado un cambio significativo en la cantidad de cancelaciones de reservas por parte de los clientes. Es un hecho de que la cancelación de forma gratuita o, preferiblemente, a un bajo costo, es beneficioso para los huéspedes del hotel, pero es un factor menos deseable y que posiblemente disminuya los ingresos para los hoteles.



El área de Data & Analytics de una empresa de reserva de hoteles busca encontrar las características en común de aquellos clientes que cancelan las reservas mediante el análisis exploratorio de datos y sus respectivas visualizaciones. Además (para más adelante) se buscará predecir que clientes podrían realizar una cancelación en la reserva solicitada a partir de algoritmos de machine learning.

El equipo de contabilidad nos ha proporcionado datos sobre todas las reservas realizadas por 36275 personas para los años 2017 y 2018.

## Objetivos y preguntas de Interés

### Objetivos

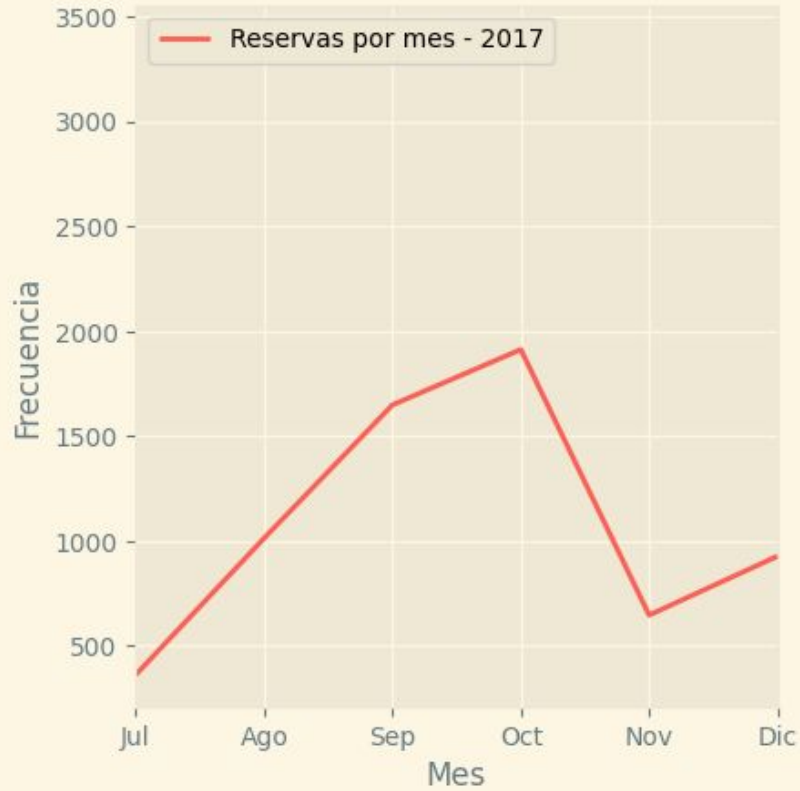
Realizar un análisis exploratorio de los datos para investigar visualmente la estructura de los datos y conocer si hay diferencia en aquellos usuarios que realizan cancelaciones en las reservas versus los usuarios que no realizan cancelaciones.

### Responder las siguientes preguntas:

1. ¿Hay un aumento de reservas con el tiempo? ¿Qué porcentaje de reservas fueron canceladas por año?
2. ¿El precio promedio por habitación es mayor en las reservas canceladas que en aquellas que se concretan?
3. ¿Se observan datos anómalos (Outliers)?
4. ¿Las reservas son en la mayoría hechas por familias o por parejas sin niños?
5. ¿Existen huéspedes que ya hayan cancelado previamente?
6. ¿Se podrían detectar que reservas podrían ser canceladas y cuáles no? Esta pregunta se responderá más adelante con los algoritmos de machine learning.

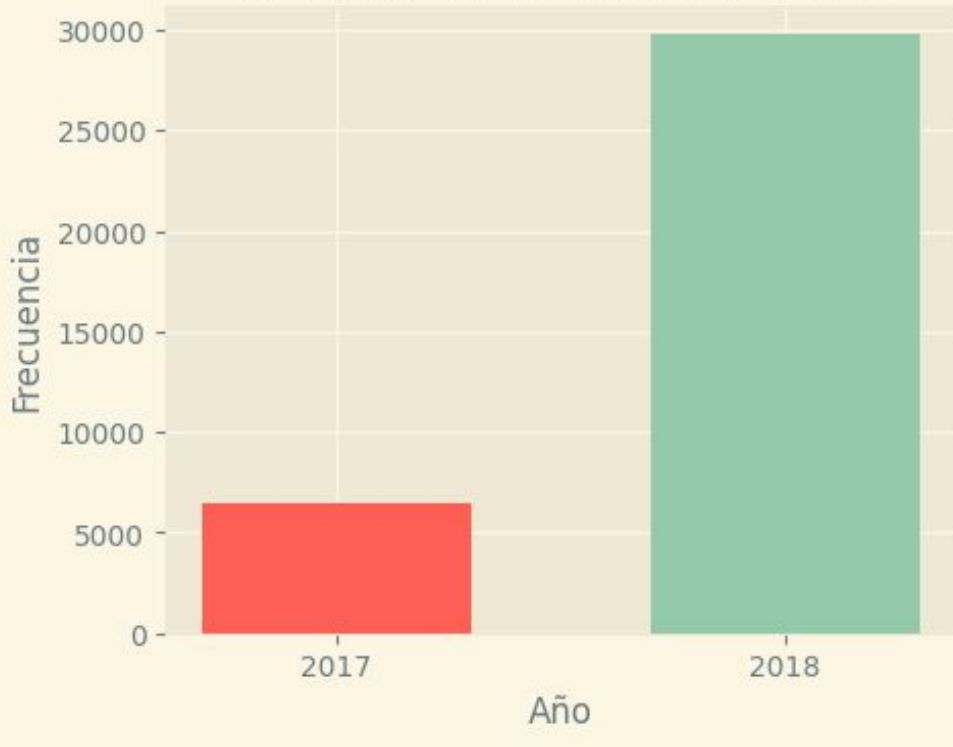
# **Análisis Exploratorio de datos**

## Número de reservas por mes (2017-2018)



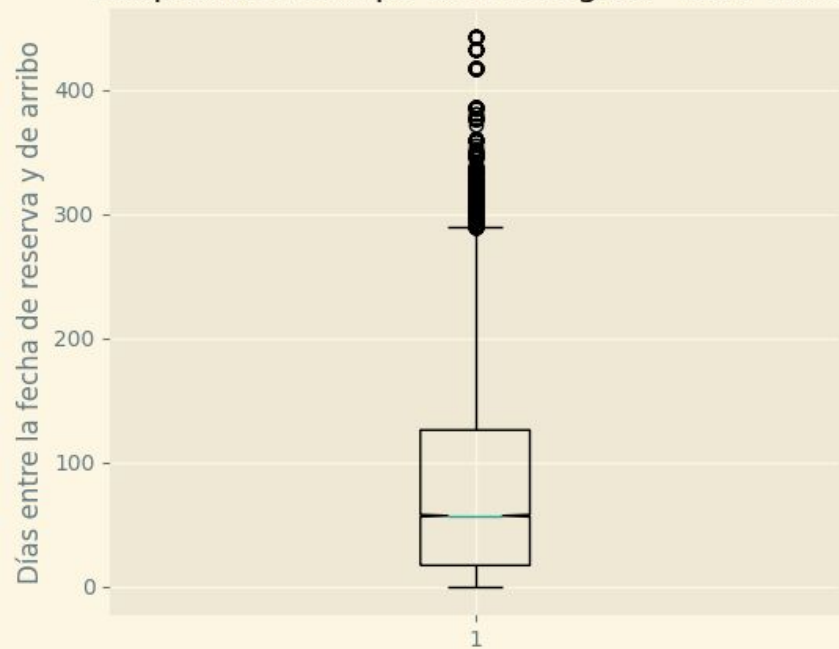
**Hay una tendencia alcista en la cantidad de reservas entre el año 2017 y el 2018**

Cantidad de reservas por año



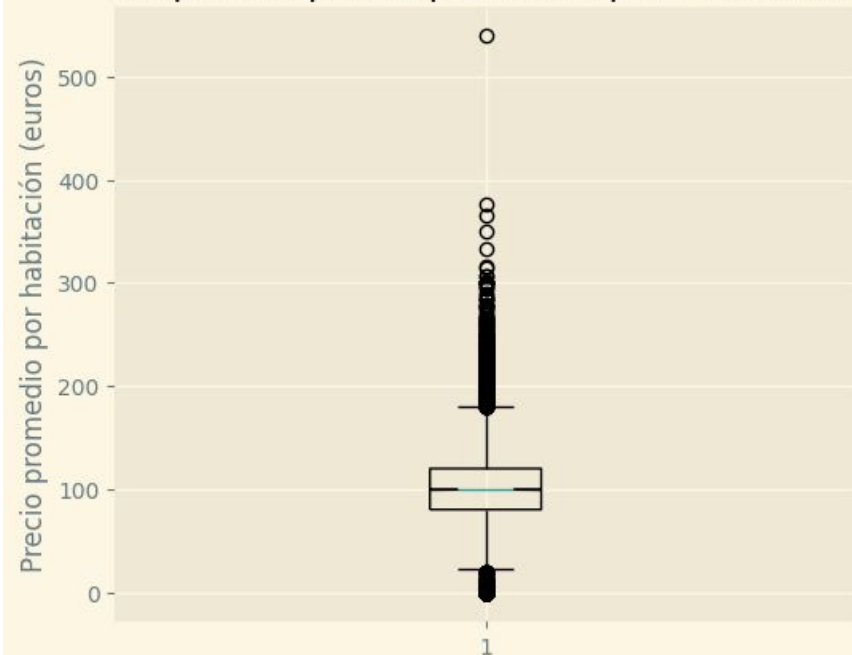
**La cantidad de reservas para el año 2018 es mayor. Una de las razones es que hay mayor cantidad de meses con datos.**

Boxplot del tiempo de entrega o "lead time"



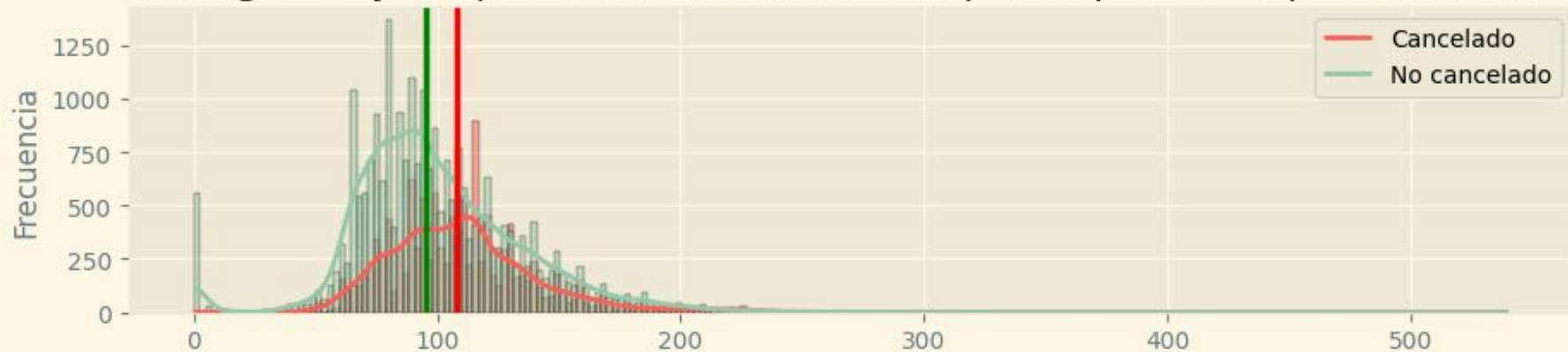
En el boxplot de reserva de tiempo de entrega o lead time se observan valores con una **mediana igual a 57 días**. El primer cuartil (25%) es 17 y el tercer cuartil 126

Boxplot de precio promedio por habitación

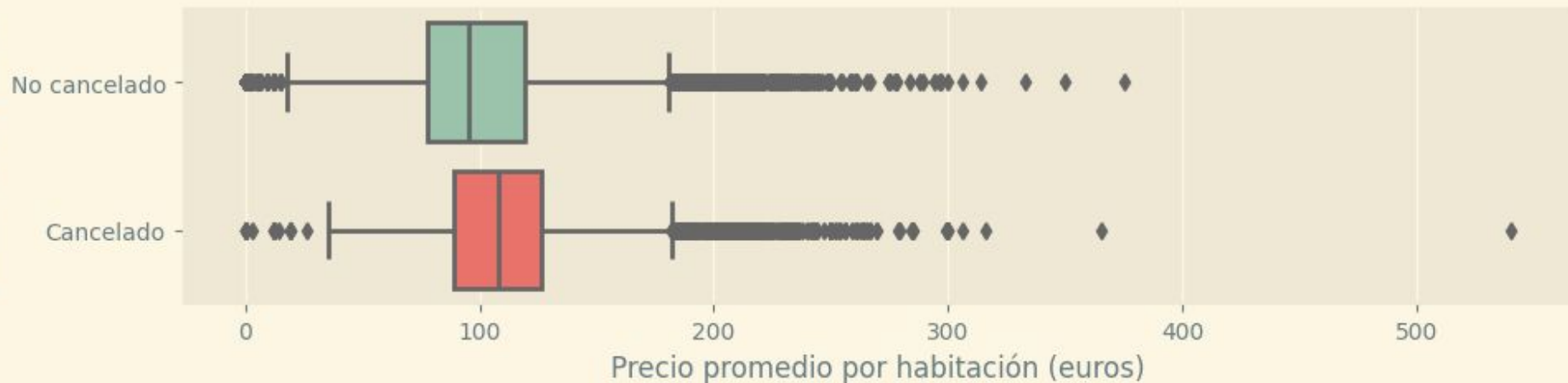


En el boxplot de precio promedio por habitación se observan precios promedios de habitación con una **mediana igual a 99.45 euros**. El primer cuartil (25%) es 80.3 euros y el tercer cuartil 120.0 euros

## Histograma y boxpot de la distribución del precio promedio por habitación



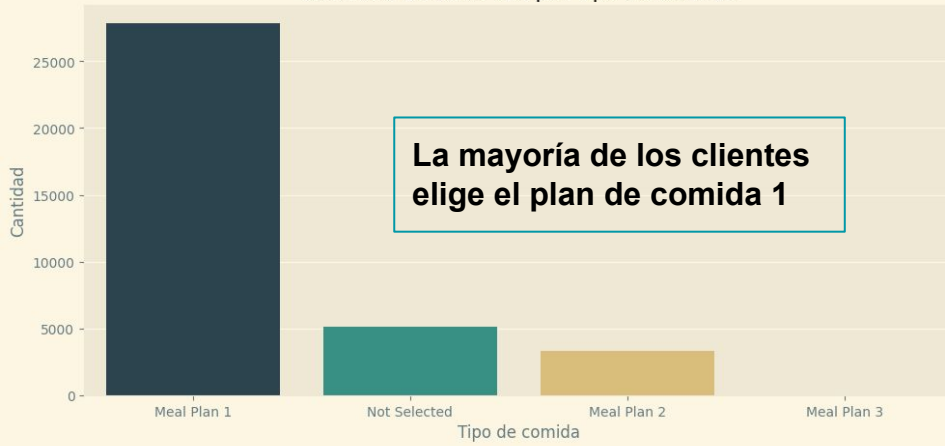
Estado de la reserva



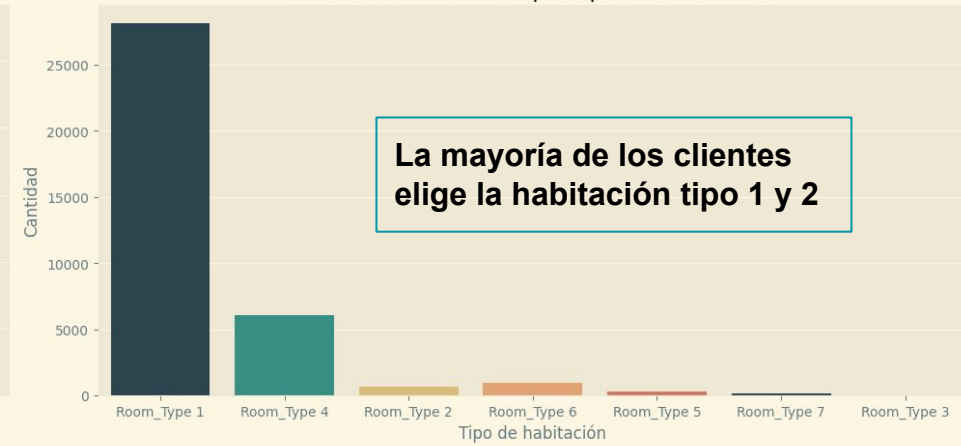
**Se observa que la mediana de los precios promedios por habitación es mayor en los casos de reservas canceladas (108 euros) versus las no canceladas (95 euros)**



Cantidad de reservas por tipo de comida



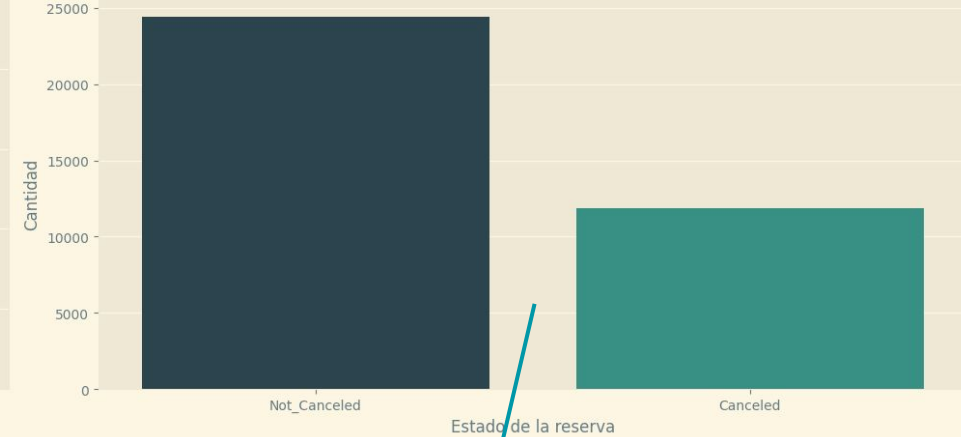
Cantidad de reservas por tipo de habitación



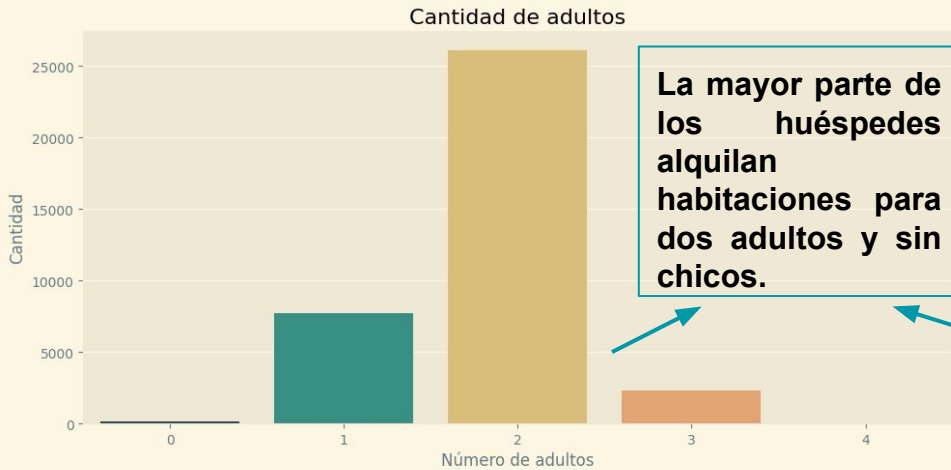
Cantidad de reservas por tipo de segmentación de mercado



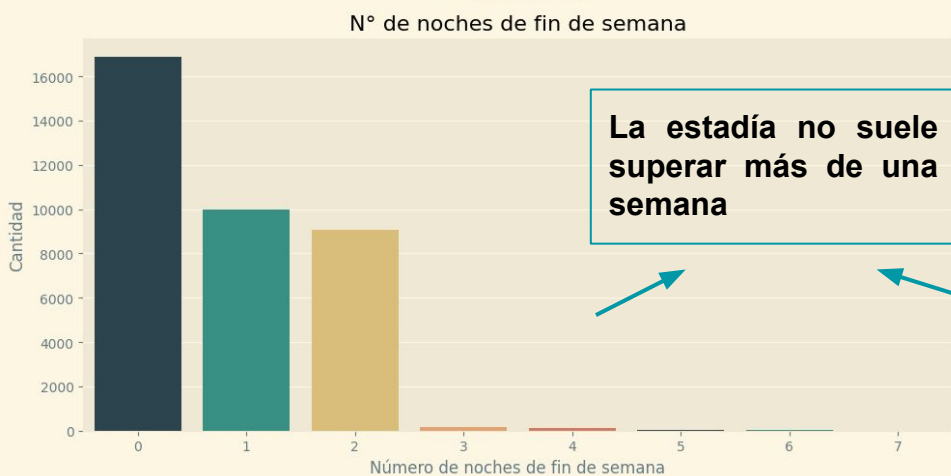
Cantidad de reservas por estado



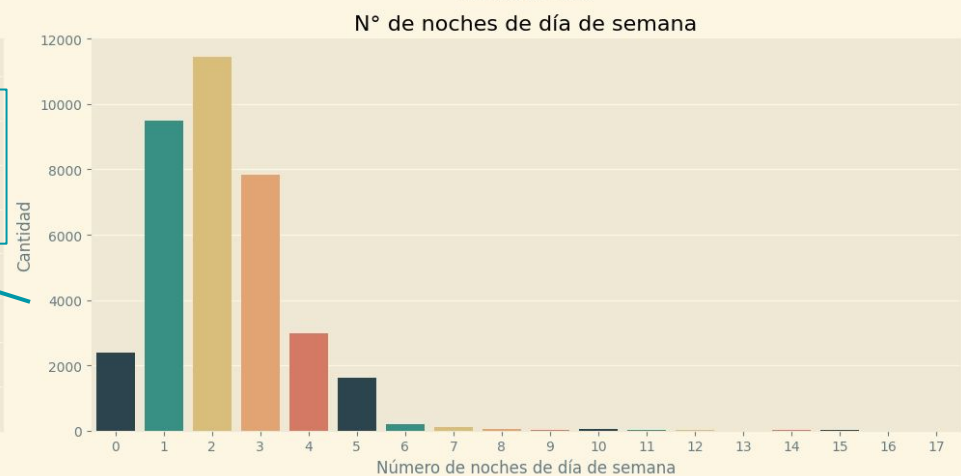
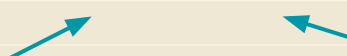
Hay muchas más reservas no canceladas que canceladas. Importante a tener en cuenta para los algoritmos de machine learning ya que los datos de la variable objetivo están desbalanceados.



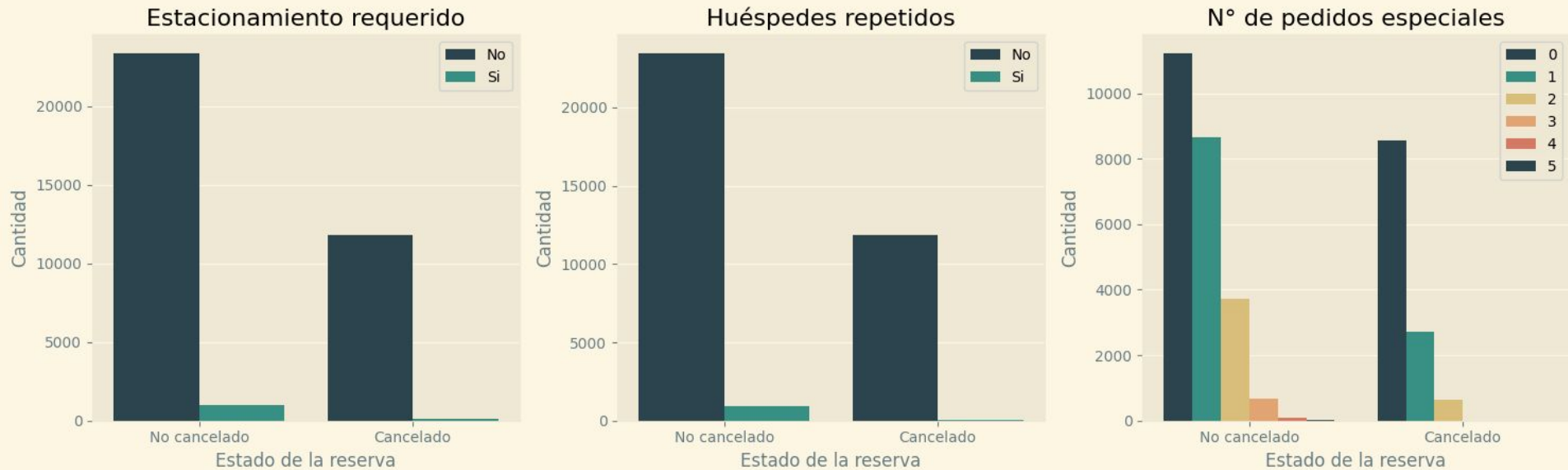
La mayor parte de los huéspedes alquilan habitaciones para dos adultos y sin chicos.



La estadía no suele superar más de una semana

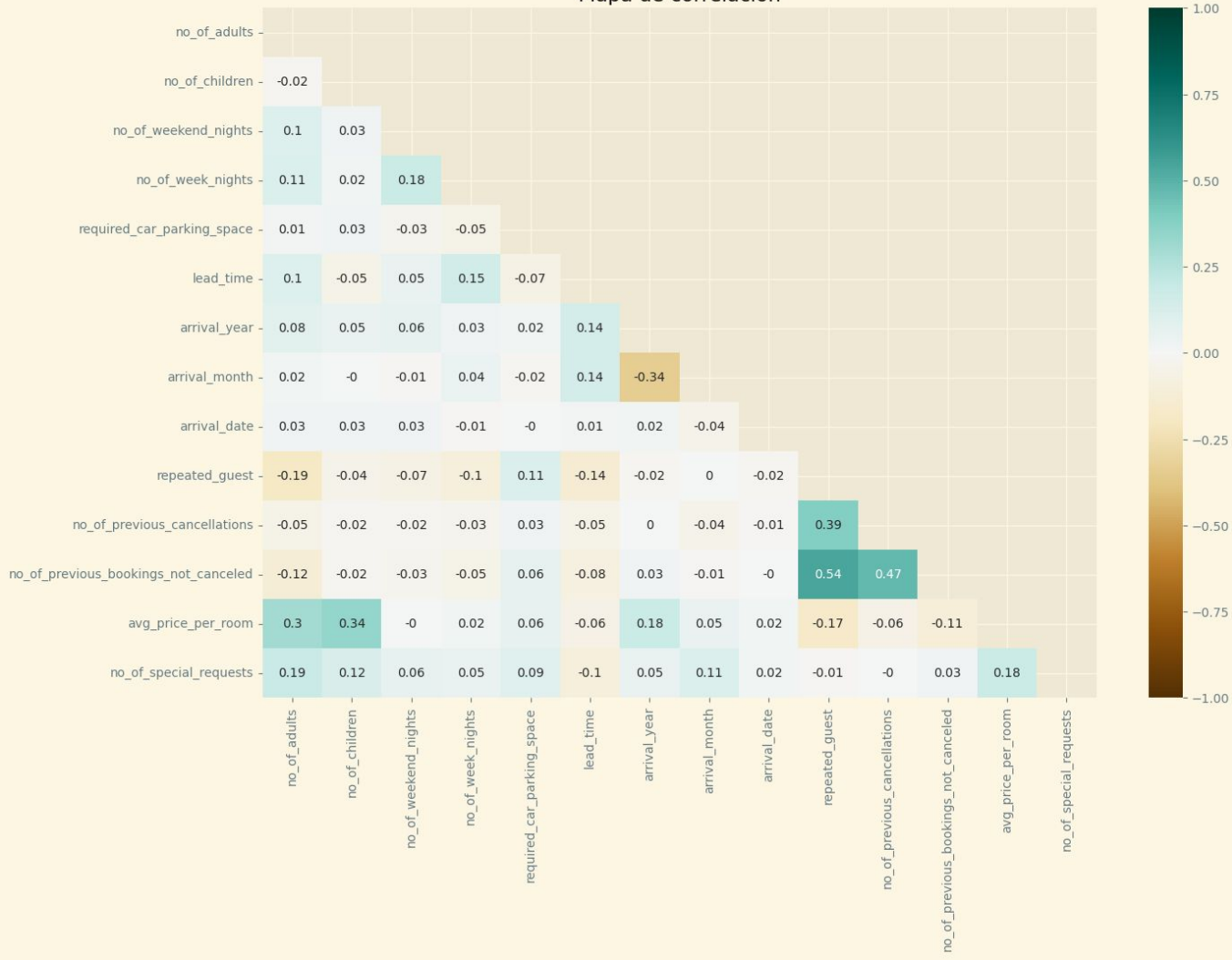


## Análisis de las variables según condición de la reserva (Cancelado - No cancelado)



**Huéspedes que solicitan un espacio de estacionamiento, o huéspedes que se han hospedado anteriormente son huéspedes que generalmente no cancelan la reserva. Además, los huéspedes que solicitan más solicitudes especiales son menos propensos a cancelar.**

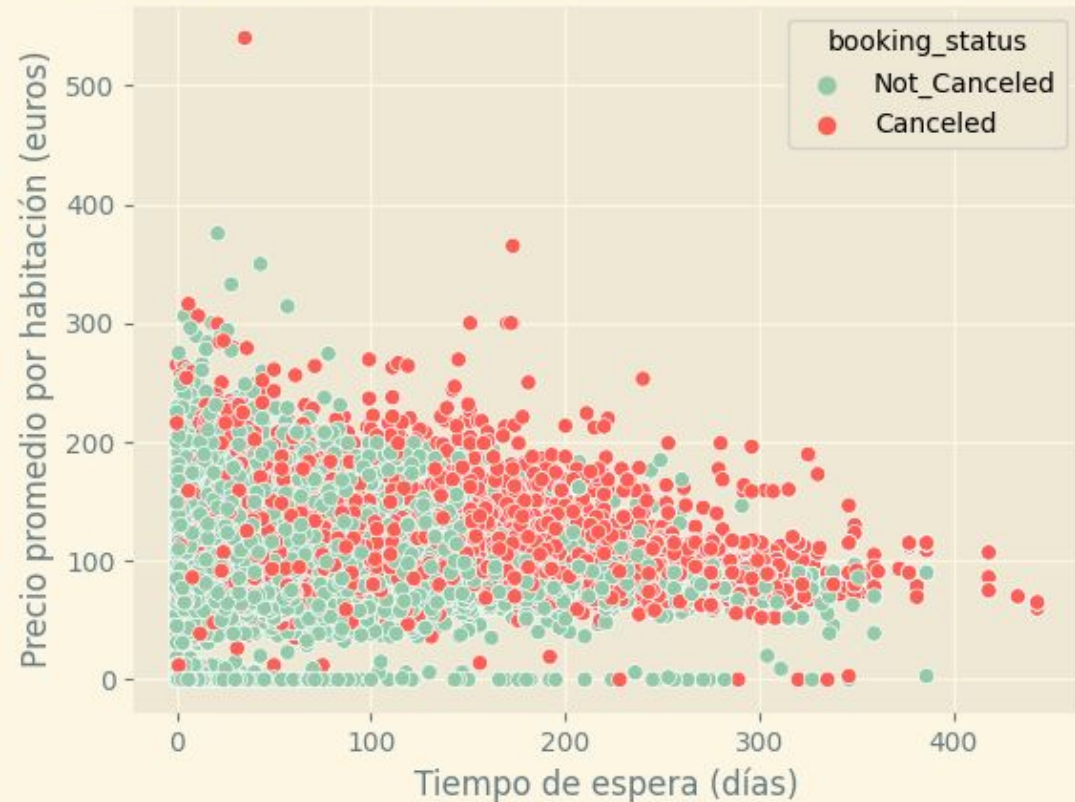
Mapa de correlación



**No hay una fuerte correlación entre las variables**

## Diagrama de dispersión

(precio promedio por habitación VS tiempo de espera)



El coeficiente de correlación entre el precio promedio por habitación y el tiempo transcurrido desde la reserva y la llegada al hotel es:  $-0.06$

↓  
No hay correlación entre  
ambas variables

Se aplicó un algoritmo de aprendizaje supervisado y de clasificación (regresión logística)

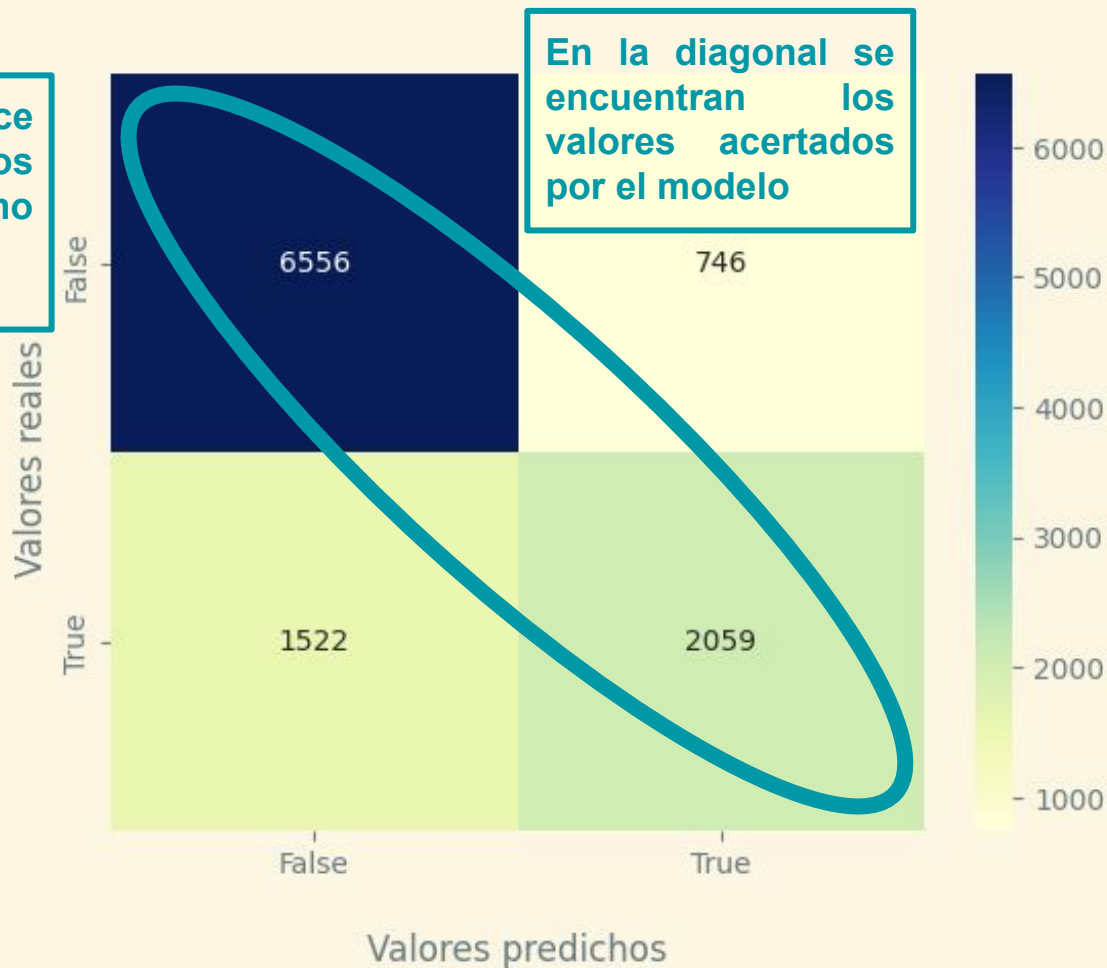
Variable etiqueta

Estado de la reserva  
Cancelado - No cancelado

- La precisión del modelo con los parámetros elegidos es de **0.79**
- La métrica F1 (más representativa para variables desbalanceadas) resultó en **0.85** para las reservas no canceladas y de **0.64** para las canceladas.

Hay un desbalance entre los casos cancelados y no cancelados

Matriz de confusión



# Insights

- La cantidad de reservas canceladas fué de 11885 (32.76%) y de 24390 (67.24%) para las reservas no canceladas para ambos años en conjunto.
- El porcentaje de reservas canceladas en el 2017 fue del 14.75 % y en el 2018 del 36.71 %, por lo que hubo un aumento en las cancelaciones del 21.96 %
- Se puede observar que en el 2018 aumentaron la cantidad de reservas para cada mes en comparación del año 2017. Además, no hay datos en los meses previos a Julio para el año 2017. El mes con mayor cantidad de reservas, tanto para 2017 y 2018 es octubre.
- En el gráfico de "Cantidad de reservas por mes", si se tiene en cuenta los años 2017 y 2018 en conjunto se puede observar que hay una tendencia alcista de reservas desde enero hasta el mes de octubre y para noviembre empieza a decaer.
- En el boxplot de precios promedios por habitación se observan valores con mediana igual a 99.45 euros. El primer cuartil (25%) es 80.3 euros y el tercer cuartil 120.0 euros. Se detectaron datos anómalos.

# Insights (cont.)

- El promedio de pedidos especiales para las reservas no canceladas es 0.76 y de 0.33 para las canceladas, indicando que si el número de solicitudes especiales aumenta, es menos probable que las personas cancelen su reserva.
- Las medidas estadísticas son mayores para los casos de las reservas canceladas que para aquellas no canceladas en el campo de precio promedio por habitación, como se observa en el histograma y en el boxplot. A modo de ejemplo, el valor para la mediana en el caso de que la reserva no haya sido cancelada es de 95.0 euros y de 108.0 euros para aquellas canceladas. Se podría confirmar con un test de hipótesis.
- Se observa que la mayoría de los usuarios eligen el tipo de comida del plan 1 y contratan habitaciones de tipo 1, siendo la reservación mayormente online.
- Complementando al punto anterior, la mayor cantidad de reservas son para dos adultos y en la mayoría de los casos sin niños.
- Como el porcentaje de los datos anómalos es relativamente chico respecto al total de registros, se decidió no excluir estos registros del análisis
- Huéspedes que solicitan un espacio de estacionamiento, o huéspedes que se han hospedado anteriormente son huéspedes que generalmente no cancelan la reserva. Además, los huéspedes que solicitan más solicitudes especiales son menos propensos a cancelar.