

Trabajo Práctico 1 : Reservas de Hotel

Informe checkpoint 1

[75.06] Organización de datos
Grupo 36

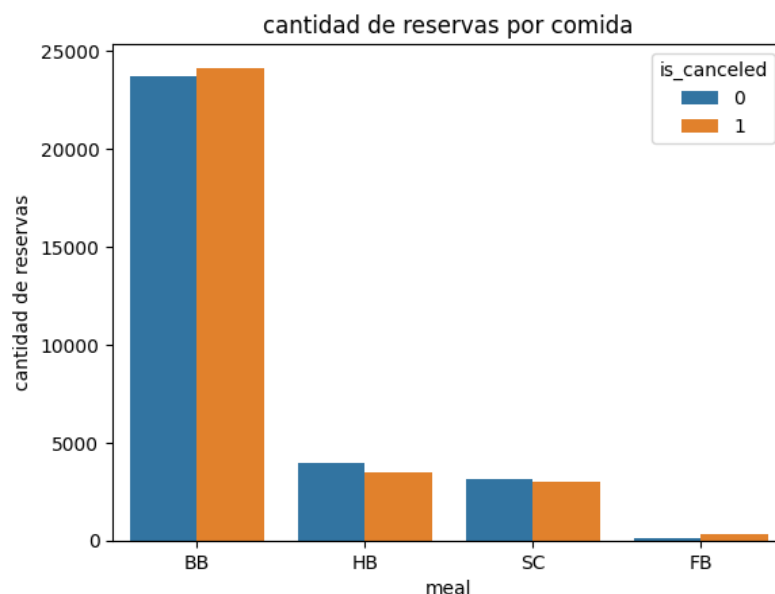
- Nicolás Tonizzo - 107820

Introducción

Se utiliza el conjunto de datos 'hotels train' para entrenar modelos que analicen los datos y puedan predecir el valor de la variable 'is canceled' de una reserva futura. Para este checkpoint se deben analizar todas las variables del dataset y proveer visualizaciones para estas, además de analizar su relación entre sí, y principalmente su relación con el target('is canceled'). También hay que analizar y trabajar los datos faltantes en el dataset.

Variables irrelevantes

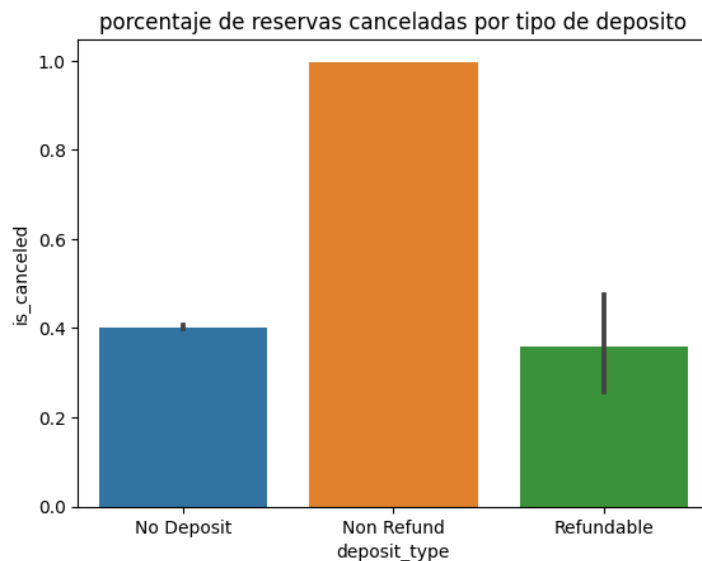
Primero de todo hay algunas variables como 'id' y 'reservation status date' que son completamente irrelevantes para el análisis. Pero también hay otras variables como 'meal', 'arrival date year', 'arrival date month' y algunas otras que en principio no tienen una relación notable con el target. A pesar de esto, no todas estas variables son descartables ya que pueden ser útiles cuando están conjuntas con otras variables.



Acá se puede ver como la variable 'meal' sola no indica mucho de sí una reserva es cancelada o no ya que todos los distintos tipos tienen aproximadamente la misma cantidad de reservas canceladas y no canceladas.

Variables relevantes

Las variables más relevantes incluyen a 'country', 'deposit type', 'lead time' y otras varias. Estas variables tienen algunos valores donde se nota una gran tendencia hacia cancelar o no cancelar una reserva. Por ejemplo en el caso de 'deposit type' vemos que cuando esta toma el valor 'non refund' hay un casi 100% de posibilidad de que la reserva haya sido cancelada.



También existe el caso de 'country' que cuando este es Portugal, las reservas suelen ser canceladas, pero la mayoría del resto de los países tienen una tendencia a no cancelar. Esto puede ser porque el hotel está ubicado en Portugal y los extranjeros harán más planeación para visitar el hotel mientras que los portugueses pueden ser más espontáneos.

