Exploración inicial

Información irrelevante:

A continuación, las variables que decidimos tomarlas como irrelevantes (no hicimos ningún gráfico que las incluya) y la justificación:

id: Porque la única información que otorgan sirve para distinguir una fila de otra.
reservation_status: Porque la variable enfoque "**is_canceled**" nos otorga información más precisa sobre lo que queremos analizar.
reservation_status_date: Porque acompaña a la variable anterior.
company: Porque casi el 95% de las reservas no especifican la compañía.

Variables cuantitativas	**Media**	**Moda**	**Mínimo**	**Máximo**
adr	102.4	62.0	-6.4	510
adults	1.9	2	0	55
children	0.1	0	0	10
babies	0	0	0	9
lead time	112.2	0	0	629
days_in_waiting_list	2.6	0	0	391

Variables cualitativas	**Posibilidades**	**Mínimo**	**Máximo**
agent	296	25.0: 1	9.0: 17004 Nulo: 7890
arrival_date_month	12	January: 3001	August: 7176
country	149	Rwanda: 1	Portugal: 27950
deposit_type	0	Refundable: 78 Non refund: 10150	No deposit: 51685
meal	5	FB: 477 Undefined: 591	BB: 47837
is_repeated_guest	2	No repetido: 60180	Repetido: 1733
is_canceled	2	Cancelado: 30941	No Cancelado: 30972

Datos faltantes:

Reservas fantasmas: Encontramos que hay 76 reservas que no tienen adultos, niños o bebés decidimos, que al ser un 0.12% de las reservas en el dataset, eliminarlas del mismo ya que su información fue probablemente un error.
Reserva gratis: Encontramos que hay 884 reservas que tienen reservas con adr con valor nulo y 1 con valor negativo, a menos que el hotel funcione a caridad, estas reservas las borraremos porque son probablemente un error, aparte al ser 1.42% dell total del dataset no deberían afectar mucho a la predicción.