**Colunas**

**hotel:** Resort ou hotel da cidade

**Is\_canceled:** Está cancelado

**Lead\_time:** tempo de espera

**arrival\_date\_year:** chegada\_data\_ano

**arrival\_date\_month:** chegada\_data\_mês

**arrival\_date\_week\_number:** chegada\_data\_semana\_número

Como já existe dados de chegada para o dia, mês e ano entendi que esta coluna poderia ser dispensada.

**arrival\_date\_day\_of\_month:** chegada\_data\_dia\_de\_mês

**stays\_in\_weekend\_nights:** número de noites de permanência em finais de semana

**stays\_in\_week\_nights:** número de noites de permanência durante a semana

**adults:** número de adultos

**children:** número de crianças

de acordo com comando “df['children'].value\_counts(normalize=True)\*100 “esta coluna apresenta 94 % dos dados com quartos sem crianças portanto, essa coluna será descartada.

**babies:** números de bebês

de acordo com comando “df['children'].value\_counts(normalize=True)\*100 “esta coluna apresenta 94 % dos dados com quartos sem crianças portanto, essa coluna será descartada.

**meal:** tipo de refeição (SC-sem refeição, BB-pousada, MP-meia pensão, FB refeição completa)

**country:** país de origem (formato ISSO 3155-3:2013)

**market\_segment:** TA-agentes de viagem, TO-operadores turísticos

**distribution\_channel:** Canal de distribuição, TA-agentes de viagem, TO-operadores turísticos

**is\_repeated\_guest:** indica se a reserve era de um hóspede repetido

**previous\_cancellations:** número de reservas anteriores cancelada pelo cilente antes da reserva atual

df['previous\_cancellations'].value\_counts(normalize=True)\*100

**previous\_bookings\_not\_canceled:** número de reservas nao canceladas pelo cliente antes da reserva atual

df['previous\_bookings\_not\_canceled'].value\_counts(normalize=True)\*100

**reserved\_room\_type:** tipo de quarto reservado

**assigned\_room\_type:** tipo de quarto atribuído a reserva

**booking\_changes:** número de alterações feitas na reserva

**deposit\_type:** tipos de reserva / com ou sem reembolso

**agent:** id da agência de viagem que fêz a reserva

em função de haver 13 % dos dados nulos, ao excluir 13 % das linhas perderia-se 13 % de todas as outras colunas portanto, optou-se por se excluir essa coluna.

**company:** Identificação da empresa / anonimato ou não

de acordo com o comando” dz = (df['company'].isnull().sum()) / len(df['hotel'])\*100 “ esta coluna apresenta 94 % dos dados nulos portanto, esses dados podem ser descartados.

**days\_in\_waiting\_list:** tempo de espera da reserva até ser confirmada

df['days\_in\_waiting\_list'].value\_counts(normalize=True)\*100

**customer\_type:** tipo de reserva, GRUPO, TRANSITÓRIO E PARTE TRANSITÓRIA

**adr:** Taxa média diária, conforme definido pela divisão da soma de todas as transações de hospedagem pelo número total de noites de estadia

**required\_car\_parking\_spaces:** número de vagas de estacionamento

de acordo com o comando “df['required\_car\_parking\_spaces'].value\_counts(normalize=True)\*100” 93 % dos dos dados indicam que nenhum estacionamento é requerido portanto, essa coluna pode ser descartada.

**total\_of\_special\_requests:** número de pedidos especiais feitos pelo cliente

São dados que teriam bastante relevância se nós tivéssemos discriminados quais são esses “requests” portanto, nesse contexto, entendo que podem sem descartados.

**reservation\_status:** Último estado da reserva, assumindo uma das três categorias: Cancelada - a reserva foi cancelada pelo cliente; Check-Out - o cliente fez o check-in, mas já partiu; Não comparecimento - o cliente não fez o check-in e informou ao hotel o motivo

reservation\_status é um tópico que diz respeito a um momento em particular, que se altera a cada instante e portanto pode ser descartado neste tipo de avaliação.

**reservation\_status\_date:** Data em que o último status foi definido. Esta variável pode ser usada em conjunto com o ReservationStatus para entender quando a reserva foi cancelada ou quando o cliente fez check-out do hotel

Este conjunto de dados está relacionado com o anterior e, portanto, pelo mesmo motivo acima pode ser descartado.