

## Datos Faltantes:

1. **Company:** Dejamos de eliminar la columna y le asignamos el valor 0 que significa "not applicable" a las observaciones con null, ya que en el paper decía que no debían ser considerados como datos faltantes, sino que eran reservas no hechas por una compañía.
2. **Children:** Decidimos no eliminar los valores nulos, y le realizamos una imputación. En este caso tomamos la decisión de imputar por la media debido a que al ser solo 4 observaciones no afecta:
  - a. La distribución de la variable
  - b. La correlación
  - c. La varianza estimada
3. **Country:** No eliminamos los valores nulls sino que ahora los agrupamos en una nueva categoría llamada 'unknown'. Estudiamos la forma de hacer imputación de tipo Hot deck o Cold deck pero no teníamos ninguna variable que nos diera una pista o relación para saber qué país podría ser, como por ejemplo hubiese podido ser una variable ciudad, coordenadas, o algo así
4. Observamos que las variables **market\_segment** y **distribution\_channel** tenían algunos valores Undefined que no habíamos visto, para ellos utilizaremos el método **hotdeck** con el cual imputaremos la variable 'distribution\_channel' en función de 'market\_segment'. En los casos donde ambos sean 'Undefined' tomamos la decisión de eliminarlos, ya que no hay otra variable la cual esté lo suficientemente relacionada para hacer una imputación.
5. **Agent:** queda igual, rellenamos con cero los nulls

## Análisis de outliers multivariada:

1. **Adults-children-babies:** Observamos que a excepción de la observación con 10 children, el más atípico de los casos era 2 adultos y 3 infantes, por lo tanto dejamos esos. Por su parte, el de 10 children decidimos imputarla con el valor 1 debido a que creemos que puede venir con un error de tipeo donde quiso agregar 1 y se le fue un cero de más
2. **stays\_in\_weekend\_nights y stays\_in\_week\_nights:** Las analizamos debido a su alta correlación pero nos dimos cuenta que los outliers que observamos en este análisis multivariados ya se eliminaron automáticamente en el análisis univariado
3. **previous\_cancellations y previous\_bookings\_not\_canceled:** Con estos observamos que la distancia de mahalanobis que nos daba más alta eran estos valores en rojo [insertar imagen] pero para nosotros los más atípicos y que se alejaban del patrón eran los que están en azul, por lo cual decidimos eliminar solo esos (9 observaciones)