

## Consideraciones para el tratamiento de datos

Para la columna Puesto de Trabajo se encontraron 869 nombres distintos. Dado esto, para la creación de la columna "Cargo" se tomaron como relevantes aquellos especificados en la hoja Cargo, y para la creación de la columna "Área" se tomaron como relevantes aquellas especificadas en la hoja Área del archivo *Formato de los datos para importación en el CRM*.

Luego crear la nueva columna área teniendo en cuenta las relevantes anteriormente mencionadas, se encontraron valores nulos, dado que en el archivo original existen nombres de áreas (dentro de la columna Puesto de trabajo) que no caen dentro de las relevantes como por ejemplo 'Junior Process Engineer', 'PIC Cessna 402 and 421C', 'Conveyors project enginner', entre otras. Para el tratamiento de los valores nulos se optó por simplemente *excluirlos* (y no eliminarlos) para generar el archivo resultante, de este modo estos datos quedan disponibles para una posterior sanitización.

### Validación de datos resultantes

Se encontró un valor erróneo de código de país para República Dominicana, el cual coincide con el código de país de USA, se optó por dejar el valor dado que el mismo se encuentra especificado en el archivo *Datos.xlsx*.

Se validaron aquellas áreas que no aparecen en el archivo resultante, para algunas de las áreas consideradas como relevantes, las cuales son: CADENA DE SUMINISTROS, PREVENCIÓN DE PÉRDIDAS, CARGA, EXPORTACIÓN, DESARROLLO ESTRATÉGICO, LEAN SIX SIGMA / BLACK BELT, no se encontraron registros en el archivo *Datos.xlsx*.