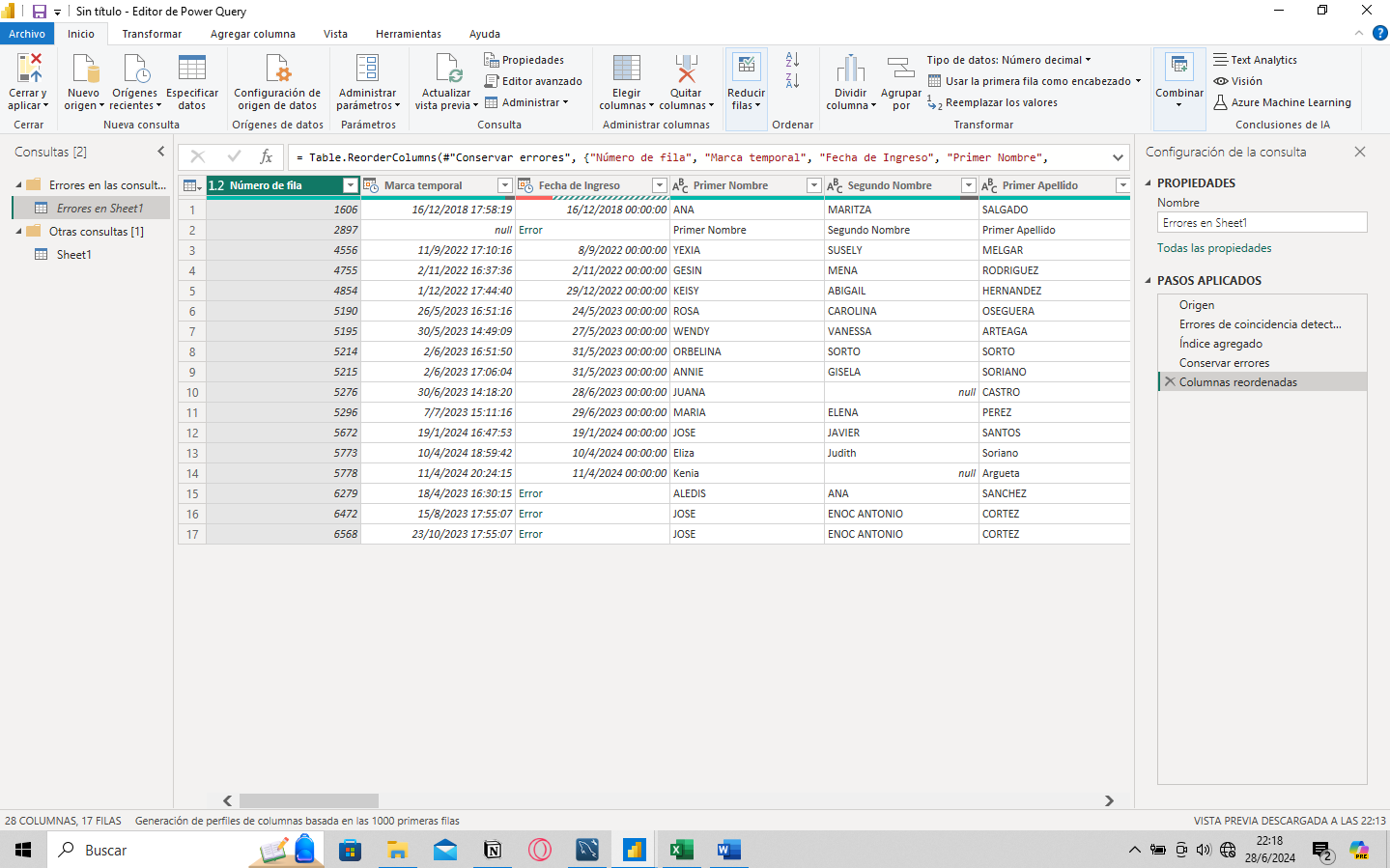
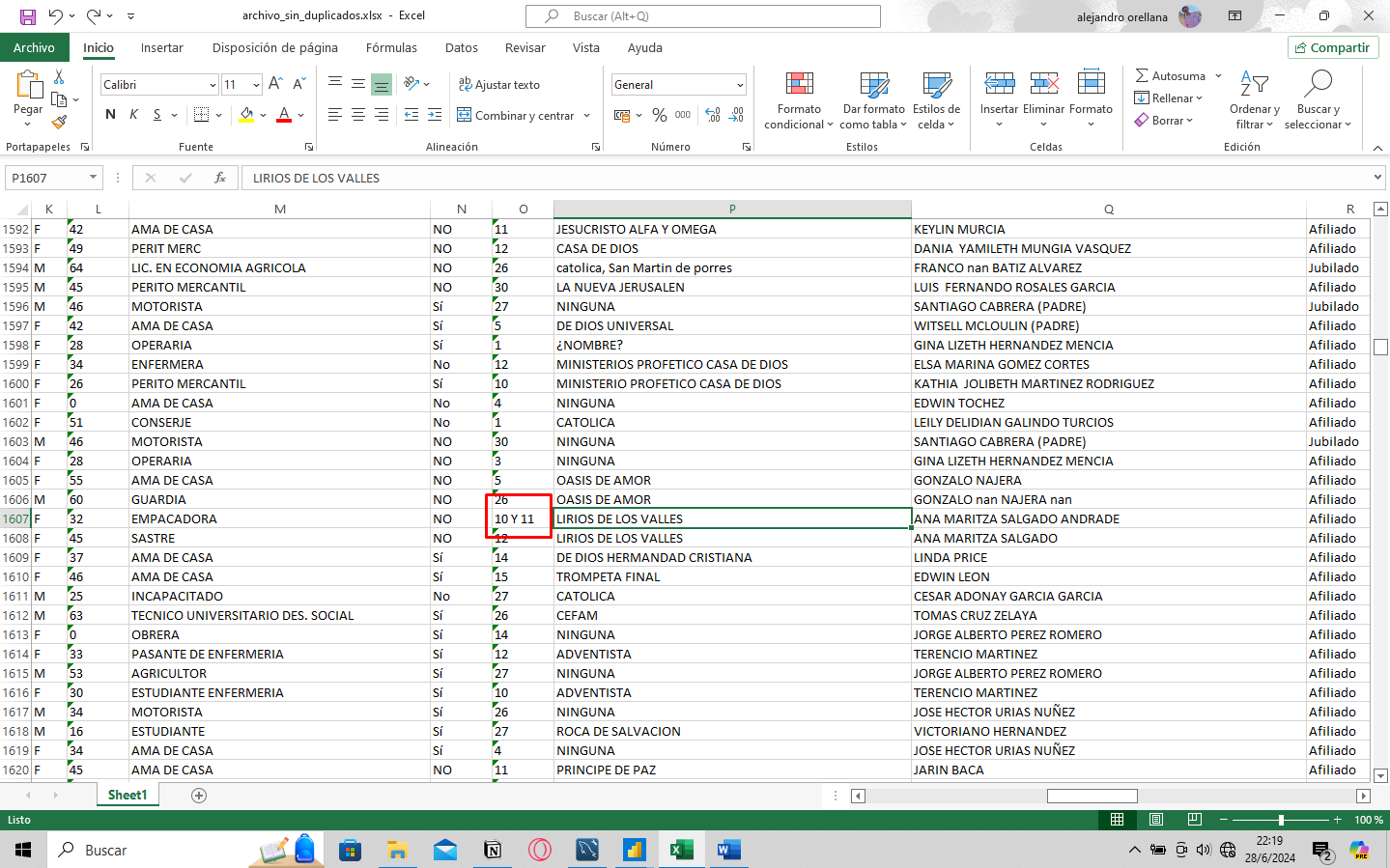
Para realizar la limpieza, utilizaremos PANDAS y POWER QUERY de POWER BI

Una vez cargamos la base de datos sin duplicados y con algunos arreglos, nos vamos a dar cuenta que POWER QUERY automáticamente nos detecta algunos errores, por ejemplo:



Estas 17 entradas, tienen errores en diferentes columnas, en este primer caso, encontramos un error en CAMA del registro numero 1606, en el que encontramos un registro de dos números, no permitido por el uso de la palabra ‘Y’

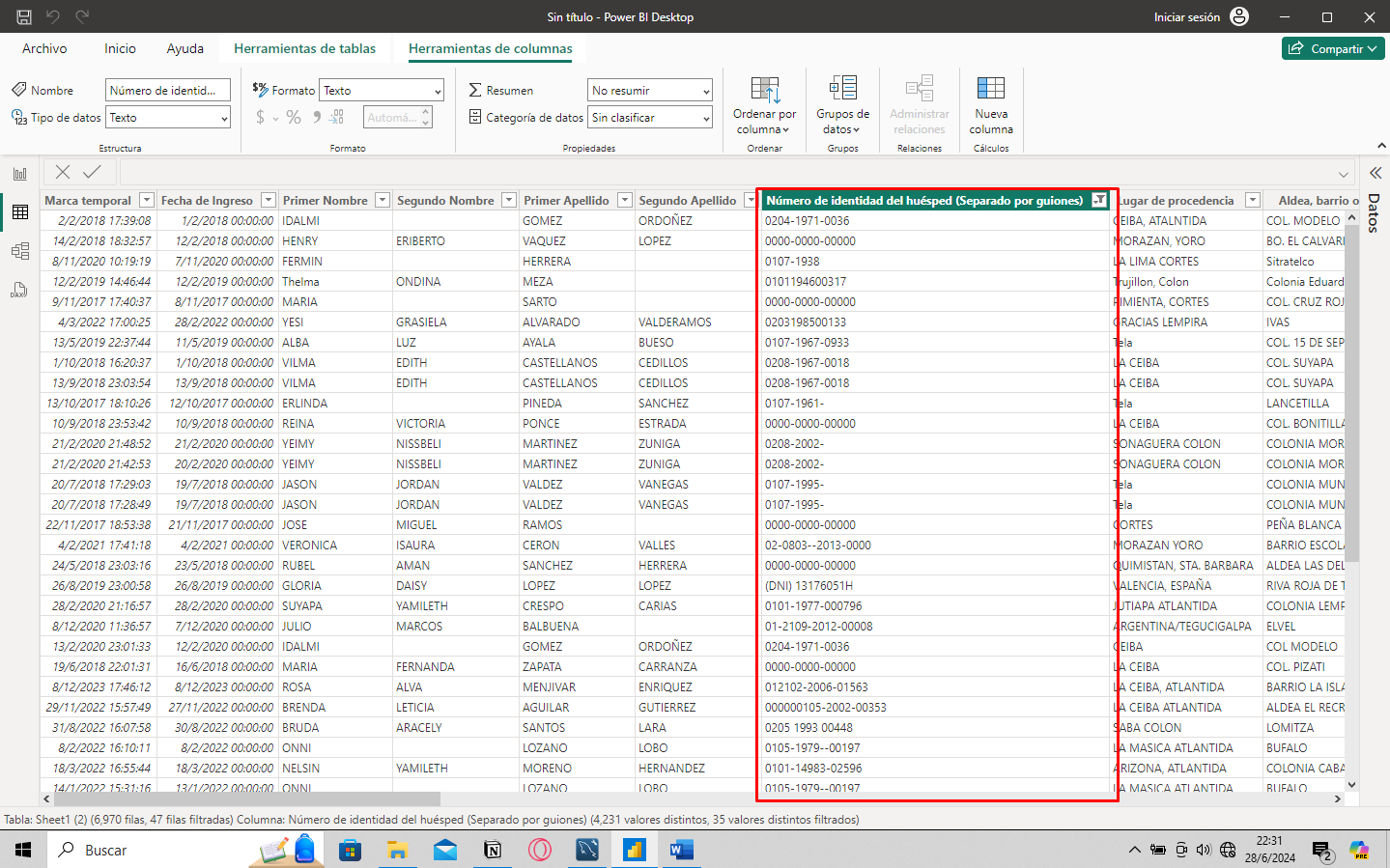


Para solucionar este problema asumimos solamente el rol del huésped de ese registro, así que dejaremos solo la cama 10

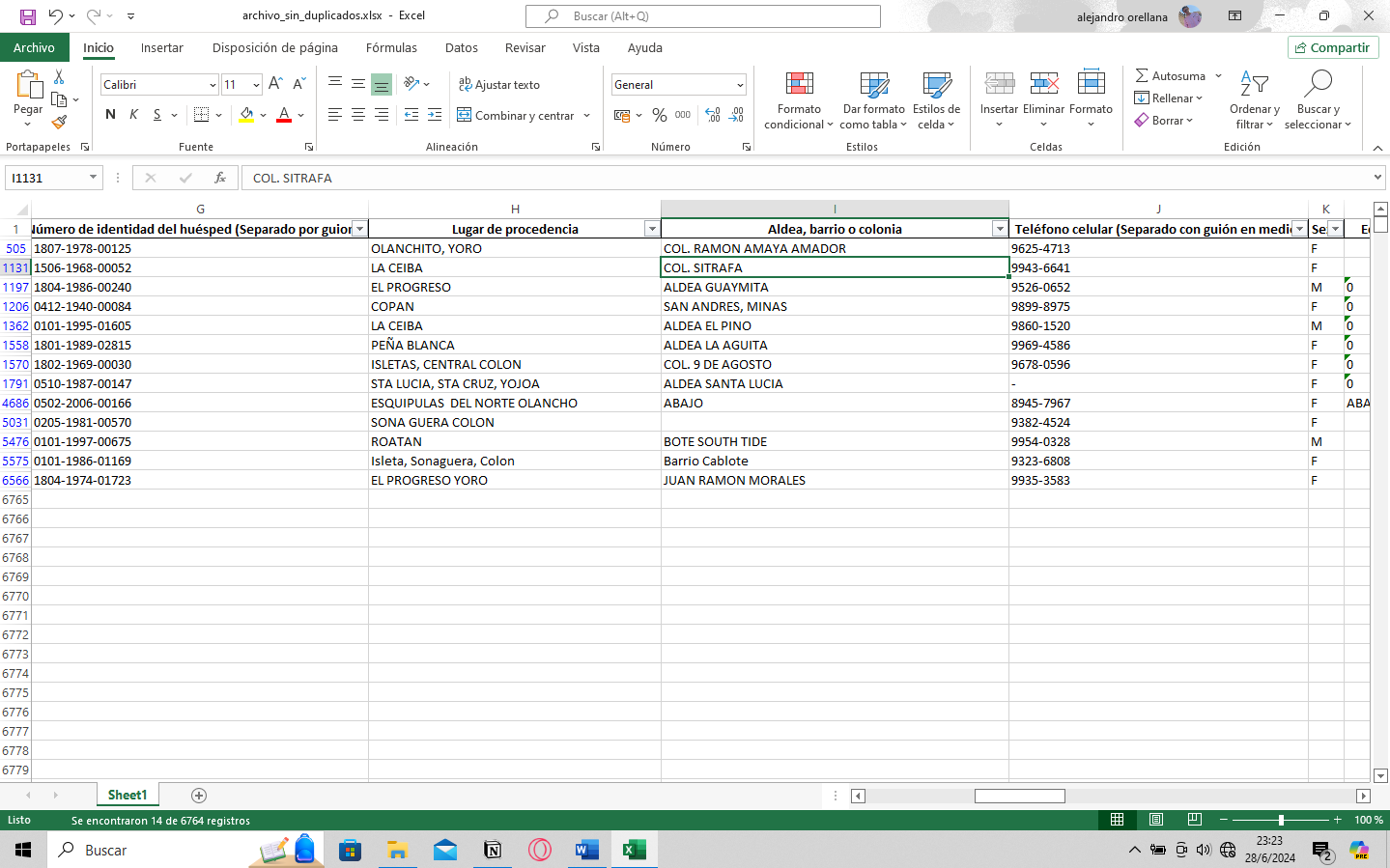
De esta misma forma, iremos solucionando poco a poco los registros que nos envían, y los que tengan errores que no se puedan arreglar o asumir, los eliminaremos.

Limpiando Números de Identidad

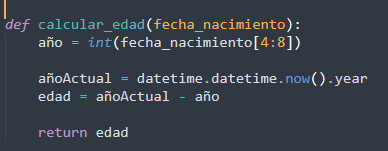
El numero de identidad es crucial, no deben de existir registros con problemas en su ID, así que, seleccionamos del filtro algunos errores, y los solucionamos, o eliminamos.



Hay algunos datos, que podemos sacar de otros, por ejemplo, el número de identidad nos puede decir la edad aproximada de la persona.



Esto, se puede hacer utilizando PANDAS, con un código que nos calcule según la identidad de la persona, una edad.

Algo así:  
  


Bien, hasta ahora, hemos afrontado problemas como por ejemplo, el numero de teléfono que contenia algunos errores minimos, que podíamos solucionar en Excel directamente, ya que eran pocos registros con un error en teléfono, sin embargo, vamos a tener que usar PANDAS de Python para algunos otros registros.